

Spring 2015

# Evolution of Prostate Specific Gene Expression Associated With Post Copulatory Sexual Selection

Scott Hergenrother

Follow this and additional works at: <https://dsc.duq.edu/etd>

---

## Recommended Citation

Hergenrother, S. (2015). Evolution of Prostate Specific Gene Expression Associated With Post Copulatory Sexual Selection (Doctoral dissertation, Duquesne University). Retrieved from <https://dsc.duq.edu/etd/649>

This Immediate Access is brought to you for free and open access by Duquesne Scholarship Collection. It has been accepted for inclusion in Electronic Theses and Dissertations by an authorized administrator of Duquesne Scholarship Collection. For more information, please contact [phillips@duq.edu](mailto:phillips@duq.edu).

EVOLUTION OF PROSTATE SPECIFIC GENE EXPRESSION ASSOCIATED WITH POST  
COPULATORY SEXUAL SELECTION

A Dissertation

Submitted to the Bayer School of Natural and Environmental Sciences

Duquesne University

In partial fulfillment of the requirements for  
the degree of Doctor of Philosophy

By

Scott D. Hergenrother

March 2015

Copyright by  
Scott D. Hergenrother

2015

EVOLUTION OF PROSTATE SPECIFIC GENE EXPRESSION ASSOCIATED WITH POST  
COPULATORY SEXUAL SELECTION

By

Scott D. Hergenrother

Approved March 30, 2015

---

Michael I. Jensen-Seaman, Ph.D.  
Associate Professor  
Department of Biological Sciences  
(Committee Chair)

---

Philip E. Auron, Ph.D.  
Professor  
Department of Biological Sciences  
(Committee Member)

---

David Lampe, Ph.D.  
Associate Professor  
Department of Biological Sciences  
(Committee Member)

---

Philip Reno, Ph.D.  
Assistant Professor  
Department of Anthropology  
Penn State University  
(Committee Member)

---

Philip Paul Reeder, Ph.D.  
Dean, Bayer School of Natural  
and Environmental Sciences

---

Joseph R. McCormick, Ph.D.  
Chair, Biological Sciences  
Associate Professor  
Department of Biological Sciences

## ABSTRACT

# EVOLUTION OF PROSTATE SPECIFIC GENE EXPRESSION ASSOCIATED WITH POST COPULATORY SEXUAL SELECTION

By

Scott D. Hergenrother

March 2015

Dissertation supervised by Michael I. Seaman, Ph.D.

Hominoid primate species differ remarkably in their social grouping and mating systems, notably including differing degrees of post-copulatory sexual selection. As the mating system of extinct hominins remains unknown and difficult to predict, it may be useful to examine more proximate phenotypes correlated with behavior. For example, chimpanzees and bonobos have a large ejaculate that coagulates into a rigid copulatory plug, presumably in response to high levels of sperm competition, while gorillas have a small semi-viscous ejaculate associated with low sperm competition. To understand the molecular basis responsible for differences in semen biochemistry among hominoid species, I completed two research projects. First, by cloning the upstream putative promoters of the chimpanzee, bonobo, human, and gorilla prostatic acid phosphatase (*ACPP*) genes into luciferase reporter vectors followed by transient transfections into a human prostate cell line, I identified the underlying nucleotide changes that reduce

expression of this protein in chimpanzee semen. Second, by mapping large deletions at the kallikrein-related peptidase (*KLK*) locus in the gorilla and gibbon genomes, I characterized the convergent gene loss and the formation of a novel chimeric gene in these monandrous species. For both the *ACPP* and *KLK* locus changes, I determined the polarity of the changes through outgroup comparison. At *ACPP*, the reduced expression in chimpanzee and bonobo is derived, and likely in response to the onset of intense sperm competition in the common ancestor of these two species. If this biochemical phenotype is indeed a proxy for mating behavior, my data provides some evidence (to be compared and contrasted with other molecular, behavioral, and paleontological data) that the last common ancestor of humans and chimpanzees was not chimp-like in its high degree of polyandry.

## DEDICATION

This dissertation is lovingly dedicated to Todd Hergenrother.

Todd,

Throughout your life, you have continuously surpassed all expectations, and you have refused to understand the word impossible. Your determination and your persistence have been and will continue to be my greatest source of inspiration.

Genuinely,

Your Brother and Life-Long Friend.

## ACKNOWLEDGEMENT

First and foremost, I would like to thank my wife, Alejandra Boza. Her love and support helped me every step of the way.

I would like to express my sincere gratitude to my adviser, Dr. Michael Jensen-Seaman, for his exceptional mentorship. Without his enthusiasm, encouragement, and guidance, this dissertation would not have been possible.

I would also like to thank my internal committee members, Dr. David Lampe and Dr. Philip Aaron. Your advice and guidance has continued to provide me with the insight needed to continue my work. You have incessantly and patiently provided me with your time, most of it unscheduled, and your support. I would especially like to thank my external committee member, Dr. Philip Reno. I am inspired by your work, and I am grateful for your time and insight.

The faculty, students, and staff in Department of Biological Sciences have continued to provide me with excellent support. My lab members Sarah Carnahan Craig, Amanda Colvin Zielen, Jennifer Vill, and Peter Chovanec have been supportive and helpful in too many ways to count. I would also like to thank the members of Auron Lab, McCormick Lab, Lampe Lab and Selcer Lab, especially Natasha Diaz and Stewart Cantlay, for technical support and advice.

Finally, I would like to thank all of my friends and family for their love, support, and encouragement, with special thanks to my amazing parents, Richard and Lynn Hergenrother, and my friends Becky Klink and Andre Furtado.



## TABLE OF CONTENTS

	Page
ABSTRACT .....	iv
DEDICATION .....	vi
ACKNOWLEDGEMENT.....	vii
LIST OF TABLES.....	xi
LIST OF FIGURES.....	xii
1 INTRODUCTION .....	1
1.1 INTRODUCTION TO THE PRIMATES .....	1
1.1.1 PRIMATE TAXONOMY .....	1
1.1.2 APE TAXONOMY AND BIOGEOGRAPHY .....	2
1.1.3 APE LIFE HISTORY .....	7
1.1.4 SEXUAL SELECTION .....	27
1.1.5 SELECTION AND <i>CIS</i> -REGULATORY CHANGES .....	34
2 EVOLUTION OF TRANSCRIPTIONAL REGULATION OF <i>ACPP</i> IN HOMINIDS.....	38
2.1 INTRODUCTION .....	38
2.1.1 <i>ACPP</i> PROTEIN STRUCTURE.....	38
2.1.2 <i>ACPP</i> ACTIVITY AND FUNCTION.....	39
2.1.3 <i>ACPP</i> GENE .....	41
2.1.4 <i>ACPP</i> REGULATION.....	42
2.1.5 DIFFERENCES BETWEEN SPECIES.....	45
2.2 MATERIALS AND METHODS .....	49

2.2.1	PCR .....	49
2.2.2	SEQUENCING .....	50
2.2.3	CONSTRUCT DESIGN .....	51
2.2.4	REPORTER ASSAYS .....	55
2.2.5	EXPERIMENTAL VECTORS USED .....	58
2.3	RESULTS .....	64
2.3.1	OPTIMIZATION AND NORMALIZATION .....	64
2.3.2	SEQUENCE DIFFERENCES AMONG THE AFRICAN APES .....	72
2.3.3	EXPRESSION DIFFERENCES AMONG CONSTRUCTS AND CONDITIONS .....	75
2.4	DISCUSSION .....	93
2.4.1	<i>ACPP</i> PATTERNS OF EXPRESSION IN HOMINIDS .....	93
2.4.2	A PHYLOGENETIC APPROACH TO <i>ACPP</i> GENE EXPRESSION PROVIDES CLUES TO ANCESTRAL MATING SYSTEMS .....	98
3	GENOMIC EVOLUTION OF <i>KLK2</i> AND <i>KLK3</i> IN HOMINIDS .....	101
3.1	INTRODUCTION .....	101
3.1.1	<i>KLK</i> FAMILY .....	101
3.1.2	<i>KLK</i> FAMILY PROTEIN ACTIVITY AND FUNCTION .....	102
3.1.3	<i>KLK</i> GENOMIC EVOLUTION .....	103
3.1.4	<i>KLK</i> ENZYMATIC ACTION AND EVOLUTION .....	104
3.1.5	<i>KLK2</i> AND <i>KLK3</i> .....	106
3.2	MATERIAL AND METHODS .....	108
3.2.1	LONG RANGE FRAGMENT AMPLIFICATION .....	108

3.2.2	SEQUENCING OF 10KB FRAGMENT .....	108
3.2.3	TRACE ARCHIVES .....	109
3.2.4	BAC LIBRARY SCREEN AND SEQUENCING.....	110
3.2.5	SEQMAN/DNASTAR.....	112
3.2.6	PROTEIN STRUCTURE PREDICTION .....	113
3.2.7	DETERMINING KLK BREAK POINT WITH DNASP.....	114
3.3	RESULTS.....	115
3.3.1	A 20KB DELETION IN THE GORILLA KLK3/KLK2 LOCUS.....	115
3.3.2	GORILLA BAC LIBRARY SCREENING AND SEQUENCING .....	116
3.3.3	CONSENSUS .....	119
3.3.4	DELETION OCCURRED BETWEEN THE FOURTH EXON OF <i>KLK3</i> AND FIFTH EXON OF <i>KLK2</i> .....	120
3.3.5	RESULTING GENE.....	122
3.4	DISCUSSION .....	126
3.4.1	SUCCESSFUL DETERMINATION OF CHIMERIC GORILLA KLK.....	126
3.4.2	GORILLAS (AND GIBBONS) USE ONE PROTEIN WHERE OTHER CATARRHINES USE TWO. 127	
3.4.3	RELATIONSHIP TO MATING SYSTEM .....	130
3.4.4	ADDING TO PUBLISHED RESULTS .....	132
4	REFERENCES.....	134
5	APPENDIX.....	147

## LIST OF TABLES

	Page
TABLE 1-1 HOMINOID BEHAVIOR AND ASSOCIATED MORPHOLOGY .....	30
TABLE 2-1 CLONE NAMES .....	63
TABLE 3-1 NCBI <i>GORILLA GORILLA</i> -WGS DATABASE TRACE FILES.....	110
TABLE 3-2 LIST OF HYBRIDIZED BAC CLONES.....	117
TABLE 5-1 ACPP PRIMERS.....	166
TABLE 5-2 GORILLA BAC HYBRID PROBES .....	167
TABLE 5-3 GORILLA CHIMERIC KLK PRIMERS.....	169

## LIST OF FIGURES

	Page
FIGURE 1-1 HOMINOID PHYLOGENY .....	3
FIGURE 1-2 THE GIBBON RANGE THROUGH SOUTHEAST ASIA.....	4
FIGURE 1-3 THE CURRENT RANGE OF ORANGUTAN.....	5
FIGURE 1-4 THE CURRENT RANGES OF GORILLA AND CHIMPANZEE.....	6
FIGURE 1-5 MATING SYSTEMS.....	29
FIGURE 1-6 LINEAR REGRESSION OF TESTES VERSUS BODY WEIGHT.....	32
FIGURE 2-1 ACPP PRIMARY AND SECONDARY STRUCTURE .....	39
FIGURE 2-2 HUMAN ACPP GENE STRUCTURE .....	41
FIGURE 2-3 ACPP REGULATION .....	44
FIGURE 2-4 DIFFERENTIAL CONCENTRATIONS OF SEMINAL DERIVED PROTEOLYTIC MOLECULES.....	46
FIGURE 2-5 WESTERN BLOT OF ACPP DERIVED FROM SEMINAL PLASMA .....	47
FIGURE 2-6 BASIC DESIGN OF PUTATIVE ACPP REGULATORY REGION INSERT .....	59
FIGURE 2-7 <i>ACPP</i> CONSTRUCT INSERTS.....	62
FIGURE 2-8 <i>HUACPP</i> EXPRESSES GREATER THAN PGL4.10.....	65
FIGURE 2-9 DAY TO DAY RAW AND NORMALIZED <i>HUΔΔACPP</i> SIGNAL.....	67
FIGURE 2-10 <i>RENILLA</i> SIGNAL.....	68
FIGURE 2-11 SIMILAR EXPRESSION OF DIFFERENT PREPARATIONS .....	69
FIGURE 2-12 SIMILAR DAY TO DAY TREND IN EXPRESSION.....	71
FIGURE 2-13 EXPRESSION WITH STEP-WISE DEACTIVATION OF FIRST EXON .....	72

FIGURE 2-14 MAPPING OF SPECIES SPECIFIC SEQUENCE DIFFERENCES.....	73
FIGURE 2-15 THE 16BP DUPLICATION .....	74
FIGURE 2-16 $\Delta\Delta ACPP$ DIFFERENCES BETWEEN SPECIES (NORMALIZED) .....	76
FIGURE 2-17 $ACPP$ DIFFERENCES BETWEEN SPECIES (RAW DATA).....	77
FIGURE 2-18 $ACPP$ DIFFERENCES BETWEEN SPECIES (RAW DATA FROM EACH DAY) .....	78
FIGURE 2-19 $ACPP$ DIFFERENCES BETWEEN SPECIES (ALTERNATIVE NORMALIZATION) .....	79
FIGURE 2-20 SYNTHETIC ANDROGEN DOWNREGULATES EXPRESSION .....	81
FIGURE 2-21 VARIABLE CONCENTRATIONS OF SYNTHETIC ANDROGEN .....	82
FIGURE 2-22 $ACPP$ DIFFERENCES BETWEEN SPECIES DRIVEN BY SYNTHETIC ANDROGEN .....	83
FIGURE 2-23 $ACPP$ CHIMERA .....	85
FIGURE 2-24 $\Delta\Delta ACPP$ CHIMERA.....	86
FIGURE 2-25 $\Delta\Delta ACPP$ TRUNCATION.....	87
FIGURE 2-26 $HU\Delta\Delta ACPP$ VERSUS $HU\Delta\Delta ACPP\_SINGLE$ .....	88
FIGURE 2-27 $HU\Delta\Delta ACPP\_SINGLE$ VERSUS $HU\Delta\Delta ACPP\_DOUBLE$ .....	90
FIGURE 2-28 COMBINED DATA FOR $HU\Delta\Delta ACPP\_SINGLE/DOUBLE/TRIPLE/QUADRUPLE$ .....	91
FIGURE 2-29 INDEPENDENT EXPERIMENTS OF $HU\Delta\Delta ACPP\_SINGLE/DOUBLE/TRIPLE/QUADRUPLE$ .....	92
FIGURE 2-30 POSSIBLE OUTCOMES OF INCREASED COPY NUMBERS OF 16BP .....	97

FIGURE 3-1 GORILLA 10KB FRAGMENT .....	116
FIGURE 3-2 BAC MAPPING TO <i>KLK</i> LOCUS OF GORILLA AND GIBBON .....	118
FIGURE 3-3 THE GORILLA CHIMERIC <i>KLK</i> .....	120
FIGURE 3-4 <i>KLK3</i> , <i>KLK2</i> , AND GORILLA CHIMERIC <i>KLK</i> DIVERGENCE. ....	121
FIGURE 3-5 HOMOLOGY BETWEEN HUMAN 30KB <i>KLK</i> REGION AND THE 10KB GORILLA CHIMERIC <i>KLK</i> REGION. ....	122
FIGURE 3-6 THE PROPOSED AMINO ACID COMPOSITION AND STRUCTURE OF THE GORILLA CHIMERIC <i>KLK</i> .....	123
FIGURE 3-7 PROTEOLYSIS IS REDUCED IN LOW AND HIGH SPERM COMPETITION CONDITIONS.....	131
FIGURE 5-1 <i>ACPP</i> PUTATIVE PROMOTER ALIGNMENTS.....	150
FIGURE 5-2 <i>ACPP</i> REGULATORY REGION GENOMIC SEQUENCE ALIGNED TO UCSC SEQUENCE.....	161
FIGURE 5-3 FINAL GORILLA <i>KLK</i> CONSENSUS.....	166
FIGURE 5-4 SIMIAN ALIGNMENT OF CONSERVED 16BP REGION IN THE FIRST INTRON OF <i>ACPP</i> .....	171

# 1 Introduction

## 1.1 Introduction to the Primates

### 1.1.1 Primate Taxonomy

The Order Primates is divided into two suborders: strepsirrhini (strepsirrhines) and haplorrhini (haplorrhines). These two suborders split approximately 60 to 70mya (Jameson *et al.* 2011). Strepsirrhines include the extant lemuriformes (lemurs) and loriformes (lorises), and the extinct adapiformes (adapiforms). The haplorrhines include the extant tarsiiiformes (tarsiers) and simiiformes (simians), and the extinct omomyiformes (omomyiforms)(Groves 2001).

Strepsirrhines are classically defined by their wet nose. All extant members of this suborder have a dental toothcomb, a grooming claw on the second digit of their feet, no post orbital closure, and a reflective layer of the eye, the *tapetum lucidum*, which helps with night vision (Nowak & Walker 1999). When compared to haplorrhines, strepsirrhines typically have a smaller brain case with larger orbits, and larger and more developed auditory and olfactory regions of the skull (Nowak and Walker 1999). The lorises can be found in Africa and Asia while the other members of the extant suborder are only found in Madagascar. Fossil evidence suggests that the adapiforms could be found throughout North America, Africa, Europe and Asia during the Eocene and Miocene (Hartwig 2002).

Haplorrhines are a much more diverse group. The tarsiers, although phylogenetically and morphologically more closely related to simians, were initially grouped with the strepsirrhines. Like the other haplorrhines, tarsiers have a dry nose as well as a fovea, a

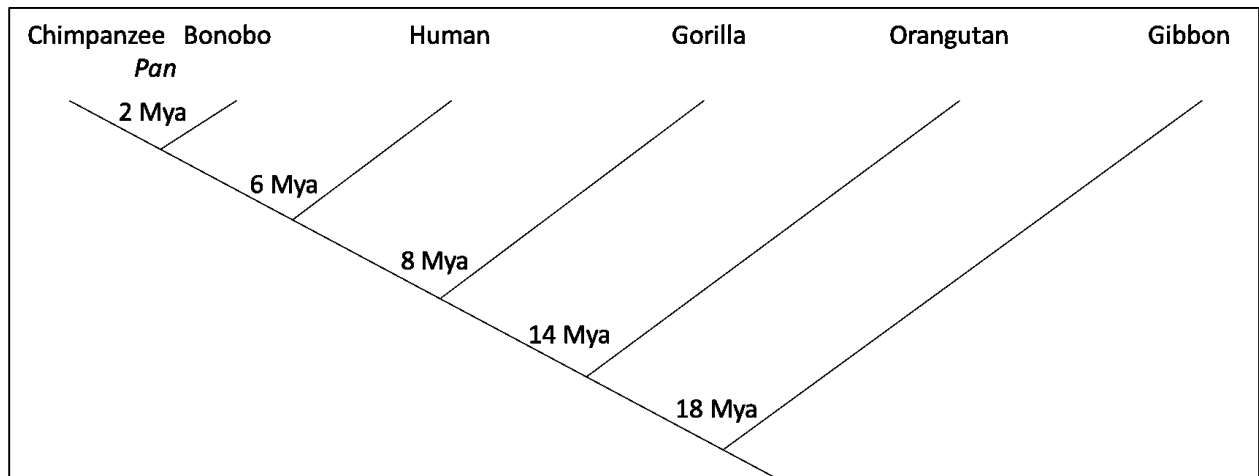


central depression in the retina that promotes sharp central vision (Nowak & Walker 1999). The simians can be further grouped into either platyrrhini (platyrrhines) or catarrhini (catarrhines) primates. The platyrrhines are also known as New World monkeys, and can be found only in the Americas. The group's most notable characteristics separating them from catarrhines are their prehensile tail, and a flattened nose with sideward facing nostrils from which their name is derived. The catarrhines, or the Old World monkeys and apes, are found in Africa and Asia, though humans belong to this group and have radiated throughout the world. Their name is derived from their downward facing nose, and unlike other primate species, catarrhines have flat finger and toe nails. Another synapomorphy uniting the catarrhines is a 2.1.2.3 dental pattern (Fleagle 1999). Omomyiforms, which are tarsier like, were found in North America, Europe, and Asia, and existed throughout the Eocene (Hartwig 2002).

### 1.1.2 Ape Taxonomy and Biogeography

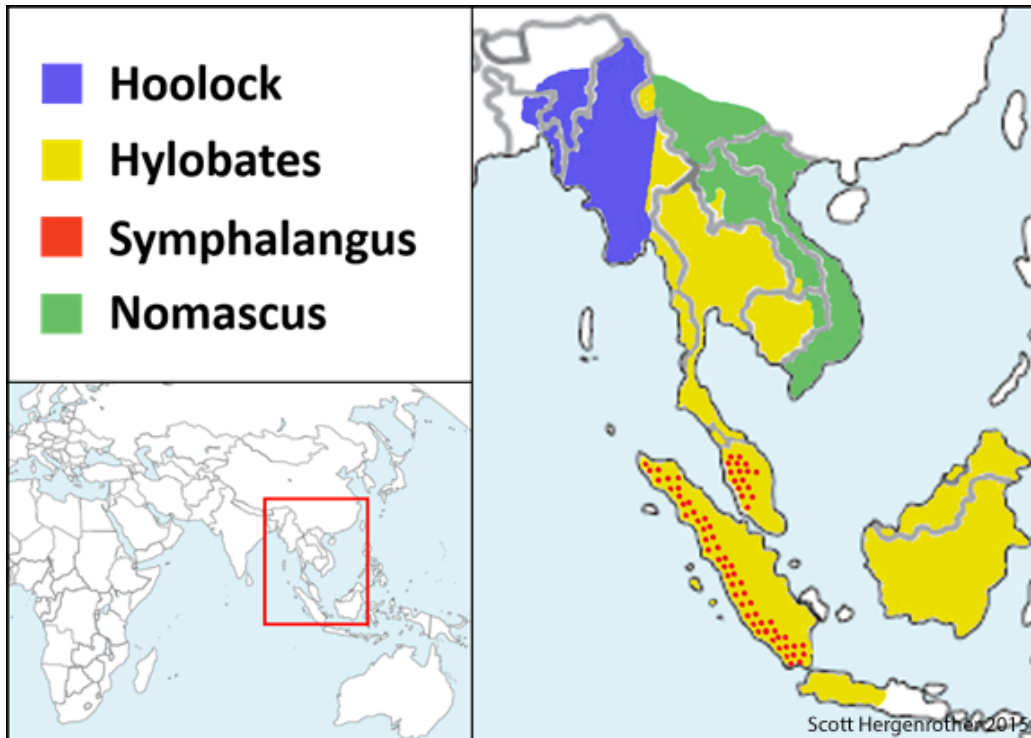
Apes lack a tail, a notable characteristic that differentiates between the Apes and Old World monkeys. Apes are divided into two families: hylobatidae (gibbons) and hominidae (hominids). Both fossil evidence (Stevens *et al.* 2013) and molecular data (Wilkinson *et al.* 2011; Springer *et al.* 2012) point to an ape split from Old World monkeys 25 to 30mya in Africa. The gibbons, or lesser apes, are currently found only in Asia (Figure 1-2). The hominids, or great apes, are found mostly in Africa except for orangutan which is found in habitats that overlap the gibbons in Southeast Asia (Figure 1-3). Along with orangutan, the

other great apes are the gorilla, chimpanzee, bonobo, and human (Figure 1-1 Hominoid Phylogeny).



**Figure 1-1 Hominoid Phylogeny**

The lesser apes, gibbons, were historically thought to be members of the same family as orangutan because of their overlapping ranges (Haeckel 1873), but because of morphological, behavioral, immunological, and more recent genetic evidence, they are now known to stem from a lineage older than that of the great apes. (Haimoff *et al.* 1982). This split between great apes and gibbon occurred about 17 to 18 Mya (Goodman *et al.* 1998; Groves 2001; Carbone *et al.* 2014). There are 14-19 species of gibbon, depending on classification structure (Mootnick 2006; Carbone *et al.* 2009), divided into 4 genera based on karyotypes: *Symphalangus* (50), *Hoolock* (2n=38), *Hylobates* (2n=44) and *Nomascus* (2n=52) (Mootnick & Groves 2005).



**Figure 1-2 The gibbon range through Southeast Asia**

Species distributions redrawn from data in Chivers (1974) and Mootnick & Groves (2005).

Orangutans (*Pongo pygmaeus*) exist solely on the islands of Borneo and Sumatra and split from the common ancestor of African apes and humans (homininae) about 14 Mya (Locke *et al.* 2011). There are two subspecies of orangutan. Borneo supports one subspecies of orangutan (*Pongo pygmaeus pygmaeus*), while the other subspecies of orangutan (*Pongo pygmaeus abelii*) is only found in Sumatra (Figure 1-3)(Meijaard *et al.* 2012). The two orangutan subspecies split around 300 thousand years ago (Kya) (Mailund *et al.* 2011) during a period of fluctuating global temperatures and drying of the region. These elements caused the sea level to drop, allowing land bridges to form between the two islands, and contracting the tropical forest habitats of the orangutan, thus separating

their populations (Verstappen 1997; Mailund *et al.* 2011). Though they are now limited to two islands, there is fossil evidence from the Pleistocene of an orangutan range that included southern China, Vietnam, Laos and Malaysia (Mitchell *et al.* 1986).



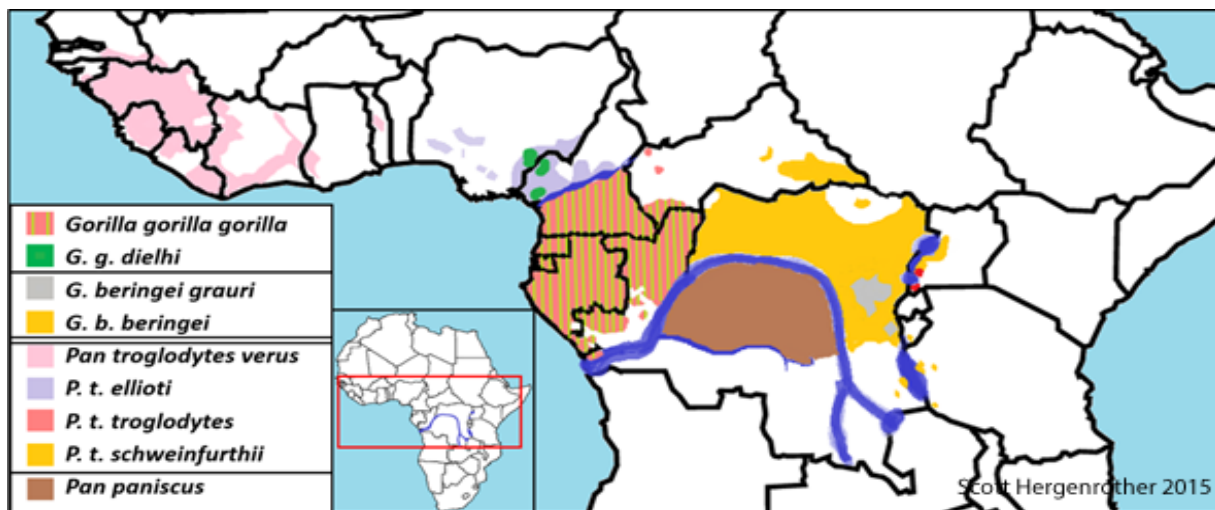
**Figure 1-3 The current range of orangutan.**

Species distributions redrawn from data in Meijaard *et al.* (2012).

The Western gorilla (*G. gorilla*) and Eastern gorilla (*G. beringei*) exist in two disparate geographical ranges north and east of the Congo River basin, separated by about 1000 km. Western gorillas are found throughout west Central Africa, as far north as southern Nigeria. Eastern Gorillas are found east of the Congo River basin to the extreme northwestern and southwestern edges of Rwanda and Uganda. Gorillas split from the human/chimpanzee lineage 7 to 9 Mya (Steiper & Young 2006; Harrison 2010; Das *et al.* 2014). As the Congo River formed, about 2 Mya, gorillas split, forming the two current species. The rise of the Congo River barrier, and the loss of lowland forest habitat that happened 2 Mya is in concordance with end of female mediated gene flow (Das *et al.* 2014) though nuclear gene

flow between the two species continued until as recent as 77 Kya (Jensen-Seaman *et al.* 2001; Thalmann *et al.* 2007). After the eastern and western gorilla split 2Mya, the eastern lowland gorilla split forming two subspecies: the eastern lowland gorilla (*G. b. graueri*) and mountain gorilla (*G. b. beringei*). The western lowland gorilla also split around the same time, forming two new subspecies: the western lowland gorilla (*G. g. gorilla*) and the Cross River gorilla (*G. g. dielhi*). (Das *et al.* 2014).

The habitats of chimpanzees (*Pan troglodytes*) and bonobos (*Pan paniscus*) sympatrically coincide with that of the gorilla in the Congo River basin (Figure 1-4). The chimpanzee lives to the north of the Congo River, and the bonobo to the south. The chimpanzee and bonobo, like the eastern and western gorilla, split about 2Mya, though there is no evidence of either male or female mediated gene flow after the split. Their common ancestors split from the hominini (hominin) 6 to 7 Mya (Steiper & Young 2006).



**Figure 1-4 The current ranges of gorilla and chimpanzee.**

Species distributions redrawn from data in Nowak & Walker (1999). Overlap between *Pan* and *Gorilla* occur, with the largest overlapping range between *G. g. gorilla* and *P. t. troglodytes*. Green stripes denoting the overlap and *G. g. gorilla* range limits.

Hominins include all species on the branch with humans after the split from the chimpanzee and bonobo common ancestor. Though humans (*Homo sapiens sapiens*) are the only extant member of the lineage, there are many different, but disputed, genera: *Ardipithecus*, *Kenyanthropus*, *Praeanthropus*, *Australopithecus*, *Paranthropus*, and *Homo*. The earliest hominin fossils, from the Pliocene and possibly as early as the Late Miocene, have been found in eastern and southern Africa (Robson & Wood 2008). While members of the genus *Homo* have roots in Africa, it is clear that some members of this genera have radiated out of Africa at multiple points, with fossils found throughout Europe and Asia (Stringer 2003; Henn et al. 2012). There is genetic evidence for African origins of modern humans, as well as at least two instances of admixture with other species of archaic human (*Neanderthal* and *Denisova*) once members of modern human had radiated out of Africa (Green et al. 2010; Reich et al. 2010).

### 1.1.3 Ape Life History

#### 1.1.3.1 Gibbon

Gibbons are a diverse family. Like the great apes and humans, they have no tail, and present a comparable dental formula. Unlike the great apes, their arms and canine teeth are relatively long, and they have sit pads like Old World monkeys. They also lack the sexual dimorphic traits found in the body, skull, and teeth of the great apes. Adult members of the genus have arms spanning in size from 0.5-1.5 m and body mass from 4-13 kg. Though male and female members of each species or subspecies within this genus are difficult to

distinguish based on size and morphological features, many species of gibbon are sexually dichromatic. Pelage changes occur mostly at sexual maturation, making it easier to identify one sex from another as well as identifying sexually mature form from sub adult (Geary 2004).

Gibbons have historically been considered monogamous because they appeared to have a mating system where both an adult male and an adult female bonded exclusively for life. An additional feature of monogamous primate groupings, like the marmoset, is that the bonded pair protects their territory from other non-related individuals. However, evidence collected in recent decades shows that the former characterization does not apply uniformly to this family. For example, extra pair copulations have been observed in the wild (Palombit 1994b; Reichard 1995) and in captivity (Reichard & Barelli 2008). Gibbons have also been recorded in groupings of more than 2 adults (Lappan 2007). Hence, gibbons' group structure may be better characterized as a *dynamic monogamy*, a system in which monogamous interactions make up a significant, but not the entirety of social interactions (Palombit 1994a). Recently, some authors (Patterson & D'Augelli 2013; Phillips 2014) have used the label *monogamish*, adapted from popular culture (Savage 2011), to refer to this type of system.

Natal dispersal is also disputed in gibbon (Shields 1982). Dispersal has been described where the offspring establish their territory on or next to the parental territory, resulting in situations where inbreeding with relatives (parents, siblings, cousins) is likely (Shields 1982, 1987). It is important to note that during these events, the dispersing male offspring have displaced neighboring territory holders and have, in some instances

obtained another males mate in the process (Brockelman *et al.* 1998). Paternal care has been recorded, but is not typical of most gibbon species (Clemens *et al.* 2008). The majority of infant care is maternal, with some sibling assistance. The behaviors that characterize an exclusively monogamous pair bonded social system do not fit these behavioral patterns found in gibbon (Bartlett 2003). It is also important to note that other than gibbon, there are no known old world primates or apes that are exclusively monogamous.

Adult pair-bonded gibbons will typically produce only one offspring every 2 to 4 years. The gestation period is on average 7 months (Carpenter 1984). A family group can have up to 4 offspring at any one time, though higher numbers of offspring are more common in captivity (Chivers 1980; Palombit 1995). Sub adult gibbons remain with their parents until 7 or 8 years of age. Sub adults as young as 4.5 years old have successfully been mated in captivity, but the average age of first observed reproduction in the wild ranges 8 to 11 years (Geissmann 1991; Reichard & Barelli 2008). There is little sign, outside of slight sexual swelling, to indicate ovulation in the gibbon (Dahl & Nadler 1992). Gibbons live 25 years on average in the wild and have about 4 to 5 offspring in this period. The gibbon life span is extended in captivity, and depending on species, they have life spans ranging from 38 to 60 years with extended reproductive periods (Geissmann *et al.* 2009).

Gibbons are diurnal and arboreal, spending most of their lives off of the ground and in the upper reaches of trees in the deciduous and evergreen rainforests of Southeast Asia. All gibbons are folivores, frugivores, and to a lesser degree insectivores (Bartlett 2007). Most gibbon species receive all of their nutrition from the trees in which they subside. As brachiation is their preferred mode of locomotion they rely heavily on contiguous forest



canopy cover. However, when moving short distances on the ground or on a tree limb they are bipedal and walk upright.

### 1.1.3.2 Orangutan

Orangutans have a dark reddish brown coat, and adults in the wild weight 30-50kg (female) and 50-90kg (male) (Rijksen 1978). They have an average head and body length of 1.25 to 1.50m, with an arm span averaging 2.25m (Markham & Groves 1990). This means that their arms are close to the ground when they are standing. Their arms, hands, and feet are very strong compared to their relatively weak legs. The orangutan's forehead is high or raised, and they do not have the pronounced brow ridge common in chimpanzee, gorilla, and the human ancestral lineage. They have a jutting jaw, with thin lips. As opposed to younger sub adult or non-dominant males, older or dominant males have flanges (cheek pads). These are deposits of subcutaneous fat that present differently in the two subspecies. The cheek pads are covered in hair and lay flat against the Sumatran orangutan's face, giving it a wide appearance, and the cheek pads of the Bornean orangutan have no hair and bulge outward. The males of both subspecies have beards and moustaches, but the Sumatran male's facial hair is thicker and fuller than the Bornean male's (Rowe 1996).

Orangutans are unique in that they are the only known diurnal primate to live in a dispersed, non-gregarious social system (Dixson 2012). Each orangutan has its own territory with the larger territories of the males defended against other males and overlapping the smaller territories of multiple females. They spend most of their adult life

alone, not including female interactions with dependent offspring. However, orangutans may come together to mate, eat, or to reinforce bonds through grooming. These temporary non-aggressive pairings are typically female-female or female-male. Male pairings are rare, and typified by violence or avoidance. Multiple adult orangutans of both sexes will sometimes come together for short periods in areas of high food density with little conflict. Multiple immature or juvenile orangutans may peacefully interact with each other and/or in the company of either adult females or males (Smuts 1987). At sexual maturity, though still socially immature, sub adult males begin to avoid interacting with adult males, and will continue this avoidance until they are able to maintain a territory of their own.

Just like their social system, the orangutan mating system is dispersed. Female ovulation is concealed, and during any period of fertility, the adult female may enter the territories of multiple dominant adult males to copulate (Schürmann & van Hooff 1986). Though female preference is for dominant adult males, they may also have non preferential forced mating “rapes” with young “vagabond” males, at the periphery or within the territory of a dominant male (Utami *et al.* 2002; Dixson 2012). These mating tactics may be successful when the young males are stronger than the females and faster than the older, slower dominant males (Setchell 2003). This leads to two different male reproductive strategies associated with secondary sexual characteristics, or two reproductive male phenotypes (Maggioncalda *et al.* 2002).

The average age of first birth is 14 to 15 years of age in the wild and slightly younger in captivity, while the first menarche occurs 1 to 4 years before first birth (Galdikas 1995; Shumaker *et al.* 2008; Knott *et al.* 2010). The female usually has one offspring, but twins do

occur. The interbirth interval is very long averaging 8 years, and it is common for an adult female to have an older juvenile offspring at time of birth (Galdikas & Wood 1990). There is no evidence for reproductive senescence in orangutans, and the average last birth occurs at about 41 years of age, with females living to an average age of 53 and males living to an average of 58 years (Wich *et al.* 2004; Shumaker *et al.* 2008).

Orangutan development includes several stages. During the first stage of development, which lasts for two years, orangutans are completely dependent on their mother. The juvenile period, from 2 to 5 years of age, is characterized by exploration and interaction with the environment in the immediate vicinity of the mother (Rijksen 1978). They are weaned at 4, just before entering the adolescent stage, but they may continue to nurse up to the age of 7. Starting around 5 years old, adolescents actively search out and group with other individuals of their own age (Munn & Fernandez 1997). They then enter a stage of sexual maturation, which begins when they are 7 to 8 years old for both males and females. During this period, female orangutans begin to show signs of sexual and social maturity. Indicators of this maturity include the philopatry establishment of an individual territory that overlaps that of their mothers, as well as sexual presentations directed towards resident males (Galdikas 1995). In contrast, male sexual maturity does not coincide with social maturity or territory establishment. During sexual maturation, males will disperse from their natal territory and enter their vagabond stage, though this may occur early during sexual maturity, or years after, when they are full grown, but still un-flanged (Galdikas 1995; Morrogh-Bernard *et al.* 2011).

Orangutans are arboreal and diurnal, spending most of their time in the trees of primary forests (Fleagle 1999). Their habitat ranges from lowland swamps, and sea level forests to mountain rainforests up to 1,500 meters above sea level. When they spend time on the ground, it is usually to get from one tree to another. Ground movement is quadrupedal, with the use of fisted knuckles of the hand for walking (Fleagle 1999). When in the trees, they use both their hands and feet for walking and climbing. During the night they sleep in nests in the trees made from surrounding foliage (Rijksen 1978). Although orangutans are mainly frugivores, they also eat foliage, mineral rich dirt, insects, eggs, and small vertebrates (Wich *et al.* 2006). They have been observed eating the carcasses of, and hunting for, larger vertebrates like the slow loris (Hardus *et al.* 2012).

### 1.1.3.3 Gorilla

Gorillas, the largest living primate, have black skin, and thick dark brown to black hair covering their body excluding their face, hands, feet, and the male chest (Rowe 1996; Nowak & Walker 1999). There are some noticeable differences between eastern and western gorillas, and even between subspecies. The hair is much longer in mountain gorillas than all other gorillas, while the western gorillas have vivid brown to red hair on their head (Rowe 1996; Nowak & Walker 1999). The eastern gorilla has a much broader chest and long face than the western. Dominant males or silverbacks have a distinct gray “silver” patch of hair on their backs and haunches, and a pronounced sagittal crest (Groves 1970). Females weigh 72-98 kg, and raised to about 1.5 meters, and males can weigh up to 181 kg in the wild and stand up to 1.75 meters. Gorillas have an arm span ranging from 2 to

2.75 meters (Miller-Schroeder 1997). Gorillas have a muzzle or snout that projects from the face with a mandibular prognathism, a lower jaw that extends past the mandible (Napier & Napier 1967). They have a pronounced brow ridge, and the males have prominent sagittal and nuchal crests (Fleagle 1999).

The gorilla's social system is typified by a single-male multi-female polygyny that includes the dependent offspring, with an average median group size of 10 weaned individuals in any park or habitat, regardless of species or subspecies (Yamagiwa *et al.* 2003). A minimum average group size in any region is always two, a silver back and female, and is consistent with new group formation (Harcourt 1978). The average group size in any one area, for any one species is not significantly different with less than 20 individuals in western lowland gorillas and 17 for eastern lowland gorillas (Yamagiwa *et al.* 2003). Larger groups of weaned individuals occur during extreme environmental conditions, like a group of 32 weaned individuals found in the abandoned village of Lossi, Congo (Bermejo 1997). Many males do not have a group, so reproductive young males not grouped with at least one female may travel together until they find a female. This occurs proportionally more often in mountain gorillas than in lowland gorillas (Yamagiwa *et al.* 2003). The single-male multi-female polygyny may also include a young reproductive male or black back, from within the group for a short period of time (Robbins *et al.* 2005). Mountain gorillas are unique in that some groups may have two related, or unrelated reproductively successful silverbacks, with a dominant silverback siring the majority, but not all, of the offspring (Bradley *et al.* 2005). These subordinate males may migrate from the group alone, in search of a mate, wait for the dominant male to die, or leave with non-related female group

member or members (Bradley *et al.* 2005). Gorillas travel with an all-male group or a male-female group. In general, gorillas do not reinforce social bonds as much as other primates, but the heterosexual groupings have less positive bond reinforcement. This reinforcement is composed mostly of dominant male-female grooming and proximity (Taylor & Goldsmith 2003). They also present more male-male and female-female aggressive interactions, when compared to the homosexual groups, which have more overall positive interactions through play, grooming and proximity (Robbins *et al.* 2005). Females with dependent young will stay with the paternal male as long as he protects his offspring from infanticide. Females without a dependent juvenile will leave the group when the male is unable to provide adequate protection due to his age, health, or harem size (Taylor & Goldsmith 2003).

Menarche begins in females around 6 years of age, but like orangutans, there is a period of infertility that lasts about 2 years (Czekala & Robbins 2001). There is very little indication of ovulation, though there is some genital swelling (Nadler 1975) that coincides with behavior changes directed at the silverback, like pursing of lips and genital presentation (Sicotte & Sicotte 2001). Females give birth about every four years (Czekala & Robbins 2001), and are the primary caregiver to the offspring during the first couple of years. Infant mortality is high, about 38% (Watts 1989) in mountain gorilla, and ranging from 22% to 65% (Robbins *et al.* 2004) in healthy western lowland gorilla populations. Weaning begins at 3, but may last up 6 years of age, when the infants become more independent. This period includes increased contact and play with other group members (Fletcher & Fletcher 2001). During this time period, the male becomes more active in

parenting, and may play with or spend time close to the younger juvenile gorillas, while actively protecting the juvenile from other aggressive group members (Stewart & Stewart 2001).

Gorillas are diurnal terrestrial quadrupeds (Fleagle 1999). All gorillas spend the majority of the day eating, resting, or traveling to another area to eat. Though they do climb trees to feed, and play with the young, the adults climb less and do not venture from the trunk of the tree when climbing. All gorillas make terrestrial nests and excluding mountain gorillas, they also make arboreal nests to sleep in at night (Fleagle 1999). Like chimpanzee, gorillas walk assisted by the finger knuckles of the two digits closest to the thumb. This leaves the hand open, allowing them to carry objects as they walk on all fours (Fleagle 1999). Eastern and western gorilla habitats are separated by about 750 km. Habitats vary within and between the two species. Eastern gorillas live in submontane and/or montane forests, with average temperatures ranging from 4°C to 15°C, and elevations from 650 meters to 4000 meters above sea level. While western gorillas live in lowland, swamp, and montane forests, with average temperatures from 20°C to 30°C, and elevations from sea level to 1600 meters above sea level (Courage *et al.* 2001). All gorilla habitats are seasonal, with at least one wet, and one dry season per year. The mountain gorilla has the most extreme habitat, with the greatest altitudes and coldest temperatures. Although all gorillas are folivores, the mountain gorilla is the most extreme with 85% of its diet consisting of leaves and the soft fleshy part of high altitude plants (Fossey & Harcourt 1977). Lowland gorillas are also mostly folivores, but a large part of their diet also consists of fruits, insects, and sometime meat in captivity (Yamagiwa *et al.* 1994; Fleagle 1999).

#### 1.1.3.4 Chimpanzee and Bonobo

Chimpanzees and bonobos (*Pan*) are all dark brown or black in skin and hair color. Unlike bonobos, chimpanzees are born with a pale face and hands that darken with age, can go bald, and have beards. Bonobos have longer hair that looks parted on the head, and keep the dark hair throughout their lives (Rowe 1996; Nowak & Walker 1999). Also, the bonobo foramen magnum is centered further under the brain case, and their ears are less prominent than those of chimpanzee (Fleagle 1999). Both species have an overall similar facial structure to gorilla, but have less pronounced snout and sagittal crest, and little to no nuchal crest. Male chimpanzees average 34-70 kg, and the females 26-50 kg in the wild (Nowak & Walker 1999). The bonobo, also known as the pygmy chimpanzee due to its less robust build, weighs about 37-61 kg in males and 27-38 kg in females (Nowak & Walker 1999). Both species stand about 0.6 to 0.9 meters tall (Rowe 1996; Fleagle 1999).

Both species of *Pan* live in a patrilineal multi-male/multi-female fission-fusion social system. This system has many adult males with high levels of male kinship, and many adult females with little to no kinship. The group itself is made of smaller dynamic multi-male/multi-female groupings or parties that leave and rejoin the larger group (Boesch 1996; Furuichi 2011). Within this type of society, each individual maintains a unique and complex set of relationships within its group. The majority of bonobo affiliative interactions are female-female then female-male (White 1996). This is in stark contrast to the male-male majority of affiliative interactions in chimpanzee with virtually no female-female affiliative behavior. In chimpanzee, however, though there is an overall increase in



female sociability during estrus, and an increase in female-female bonding when lactating (Goodall 1986; Wrangham *et al.* 1996; Pepper *et al.* 1999). Whereas bonobo group acceptance and rank acquisition is determined by the alpha female (Waal & Lanting 1997), in chimpanzee group the same feature is determined by a clear male linear dominance hierarchy (Goldberg & Wrangham 1997). This is a hierarchy that has a distinct but dynamic chain of command in which there is only one dominant (alpha) male, with the next individual in the hierarchy, the (beta) male, being dominant to all but the alpha male. The bottom of this hierarchal system is populated with young non-kin females and their offspring (Goldberg & Wrangham 1997).

Bonobo females may experience menarche between 6 and 11 years of age (Vervaecke *et al.* 1999), while chimpanzees females experience it between 8 and 11 years of age (Atsalis & Videan 2009). Though bonobos first menarche is earlier than chimpanzee, both species experience a period of infertility coinciding with female natal dispersal, and have their first offspring at about 13 years of age (Vervaecke *et al.* 1999; Atsalis & Videan 2009). Sexual swellings appear for a longer period in bonobo than in chimpanzee, with a pre-swelling, swelling, post-swelling, and menses stage, and an ambiguous period of peak fertility (Thompson-Handler *et al.* 1984). This ambiguity is thought to be part of the adaptive processes that has led to increased promiscuity and decreased intra-group aggression including infanticide in bonobos (Waal & Lanting 1997). Though mating hierarchy and mating strategies differ by genders in each species, in both of them a female will mate with multiple males, multiple times, during any given period of estrus (Waal & Lanting 1997). In chimpanzees, the hierarchy is determined by dominance, with the more dominant males

copulating more than the less dominant males, particularly during the periods of greatest fertility (Goodall 1986; Whiten *et al.* 1999). Chimpanzee females may also secretly, forced or willingly, leave the group with another male or possibly to mate with males from neighboring groups (Gagneux *et al.* 1999; but see Vigilant *et al.* 2001). Though these behaviors increase the chance of reproductive success, they may also increase the chance of male mediated infanticide (Nishida & Kawanaka 1985; Gagneux *et al.* 1999). In bonobos, promiscuous sex, though ultimately resulting in reproduction, is based on supporting social organization, hierarchy and cohesion, as well as facilitating stress reduction. Along with sexual swellings, promiscuous sex occurs independent of estrus (Waal & Lanting 1997).

In both species, the interbirth interval is about 4 to 6 years, with a life span of about 40 years in the wild. However, but some chimpanzees have lived much longer in captivity. Offspring are cared for almost exclusively by the birthing mother, though siblings also assist in care. Bonobos have a slower rate of development, and may have increased maternal care through the developmental period (Kuroda 1989). Both species are weaned between 4 and 6 years of age. Sub adult, post menarche female offspring of both species begin to have decreased interactions with their mother until dispersal. As chimpanzee males become reproductive, social bonding and interactions with other males in the group become more important than kin relations, but the bonobo male rank is connected directly to the mother's rank in the group (Goodall 1986; Kuroda 1989; White 1996; Boesch *et al.* 2002). This is characterized by lifelong within group maternal and kin affiliation (White 1996).

Chimpanzees and bonobos are quadrupedal knuckle walkers. They, like gorillas, walk with open hands (Fleagle 1999). Both also move via limited brachiation, and bipedalism, but the bonobo is the most adapted for bipedalism, with better weight distribution due to the position of the foramen magnum and long thigh and foot bones. Although both species are somewhat arboreal, the bonobo spends more time traveling in the trees (Doran 1996). Chimpanzees and bonobos build individual arboreal nests to sleep in at night. Both species' diets are mainly frugivorous, though they supplement their diets with foliage, nuts, seeds, insects, flowers, bark, soil, eggs, honey, and meat (Goodall 1986; White 1996). The acquisition of some of these foods, such as honey, nuts and insects, is often improved by the use of tools (Goodall 1986; Gruber *et al.* 2010). Although both species are opportunistic meat eaters, chimpanzees will form complex hunting parties to track down mammalian prey including monkeys and warthogs (Goodall 1986).

#### 1.1.3.5 Modern Human

Modern humans (*Homo sapiens sapiens*) have become the most numerous single species of extant ape on the planet with the current population exceeding 7 billion. Humans are unique from other apes with large upright bodies, long legs, large brains, increased meat eating, and unique early and late life histories (Anton *et al.* 2014). Humans include individuals and subpopulations with a wide variety of morphological, social, behavior, and ecological traits. Hair color in human populations ranges between shades of black, red, brown, and white, and skin color ranging between multiple shades of brown. Men have characteristic beards, but beard density and coverage varies greatly. Although hair covers

the majority of the human body, most of it is shorter, more fragile, less pigmented, and less dense than in other primates. This has led to the humans being described as hairless or as the “naked ape” (Newman 1970; Pagel & Bodmer 2003). Our skeletal features and dentition are highly similar to other great apes, with a couple of notable differences. Human canine tooth size is greatly reduced, our premolars are wider, and many times, the third molar is absent or reduced in size (Fleagle 1999). The brain case is larger, with an enlarged cranium lacking well defined brow ridges, and underdeveloped crests. The foramen magnum is located directly under the skull, and the jaw does not extend out as far as in the other apes (Fleagle 1999). Also, the human features associated with walking (the short wide hips, long leg bones, long heel bones, long metatarsal bones, and short tarsal bones) are unique among apes. Humans are the only primate to have a fixed hallux, leading to loss of the opposable thumb on the foot. Human height and weight vary, but on average humans are about 1.6 to 1.75 meters and 47 to 78 kg in men and 42 to 73 kg in women. Men, on average, are 1.1 to 1.2 times heavier than women, and 1.06 times taller (Fleagle 1999; Dixson 2009).

Humans have been typically characterized as monandrous, living in social systems that are either monogamous or polygynous (Darwin 1871), though polyandrous human systems do exist (Fleagle 1999; Starkweather & Hames 2012). Humans, when compared to all other primates, have the most diversity in their social organization (Fleagle 1999). Although pre-Neolithic humans were hunter gatherers, the current human populations live in communities that range from nomadic to sedentary, and societies that range from hunter gatherer to agrarian with >95% of modern humans living in sedentary and agrarian

societies (Harding 1982). Humans are unique in their ability to adapt to new social settings. They not only peacefully surround themselves by strangers on a regular basis, but also engage in hyper-cooperative behaviors, or behaviors that sacrificially benefit others (Hrdy 2009; Tomasello & Vaish 2013). Importantly, humans of both genders participate in dispersal, or fission fusion, in many instances they separate themselves from their mates and offspring to participate in activities with strangers of both genders before returning to their familial setting (Aureli *et al.* 2008; Hrdy 2009). Humans can live in many social settings, with any number of individuals, related and/or nonrelated, with one or both genders of any age. However, there are cultural constraints specific to each group that determine how humans interact within those cultures, and how they think about their interactions (Fleagle 1999; Peterson 1999; Costa *et al.* 2001; Ozer *et al.* 2013). Like the other primates, these qualities makes it difficult to categorize humans into any one social structure (Fleagle 1999). Categorizing the human social system using biological metrics or models from other species is also difficult. Using relative testis weight suggests that humans are polygynous, and are more polyandrous than gorillas (Harcourt *et al.* 1981; Dixson 2009). Using sperm midpiece volume, and mitochondrial density, humans look more monandrous than chimpanzee (Anderson & Dixson 2002; Anderson *et al.* 2007).

Well-developed step wise models, assuming high levels of mate competition, or a chimpanzee-like human ancestral condition have been proposed. These models include the use of anatomical and behavioral correlates to explain the steps that would be necessary for a transition from polyandry in our ancestors towards a current system of increased or “strong” pair bonding. These models include a transition from promiscuity towards

increased pair-bonded monogamy driven by female choice and male provisioning, concealed ovulation, and greater paternal care (Lovejoy 2009; Gavrilets 2012). Conversely, other step wise models, using similar anatomical and behavioral correlates, have been proposed using a starting point that more closely resembles the gorilla monandrous system. The shift towards the modern human system is explained using similar analyses as before, but by reduced aggression among males within groups, then reduced aggression between groups with a simultaneously shift towards bonding with a single mate (Nakahashi & Horiuchi 2012; Chapais 2013).

Human females experience menarche on average at about 9-13 years of age, and unlike other primates, will go through reproductive senescence at about 49 years of age, and then continue to live many more decades as a post reproductive individual. Human male fertility develops during the same time frame as females, with culturally specific and culturally obfuscated periods of reproductive senescence and parental investment. Female natal dispersal in humans is estimated to be about 67% in multicultural studies(Hrdy 2000), and 56% in only foraging societies, or in societies that are assumed to mimic the ancestral condition (Hrdy 2000). Though different human societies have describable natal dispersal, there is no real gender-specific trend, with humans of both genders dispersing as well as maintaining familial relationships throughout life.

The average human new born weighs 3.25 kg. Like other apes, humans usually have only one offspring per birth, but having more than one does occur. Compared to other apes, humans are larger and more helpless at birth (Fleagle 1999). The average number of offspring depends heavily on culture, with developed nations like the United States, China,

and the European Union having low birthrates at or below the population replacement rate of 2 offspring per woman, and under developed nations having birth rates as high as 5.4 children per woman in Afghanistan, and 6 children per women in Somalia (United States. Central Intelligence Agency. 2014).

The ancestry of modern humans can be traced back about 200 kya (McDougall *et al.* 2005). They dispersed from Africa as early as 72 kya, and had populated every continent excluding Antarctica by about 12.5 kya (Oppenheimer 2012). Currently, humans continuously occupy every continent on earth. Also, they have had long term habitats in regions that are not hospitable to human life like those in earth's orbit, Mir and the International space station, as well as underwater habitats like Conshelf II and SEALAB I and II. This is due, in part, to the unique human ability to dexterously and intensely manipulate their surrounding environment to one which fits their needs. These abilities are enhanced by human bipedal, or upright walking, which allows humans to uniquely manipulate objects. These unique abilities range from reshaping the land to form areas for agricultural and mining to the production of goods produced from these areas. Humans in general are omnivores, and though culture and environment have a huge effect on diet, humans consume fruits, vegetables, grains, meat and eggs. Some human populations contain adaptations related to their specific diets like the ability to digest lactose (lactose persistence) after weaning in populations reliant on milk (Holden & Mace 1997), and increased gene copy number and subsequent expression of alpha-amylase salivary starch digestion enzyme in populations that have been more starch reliant (Perry *et al.* 2007).

#### 1.1.3.6 Ancient Human

Using comparative data obtained from the environmental, fossil, molecular, and archeological record, some testable models can be produced to infer the life histories of ancient hominins (Anton *et al.* 2014). Ancient humans have an exceptionally large fossil record, which may allow us to elucidate a more complete ancestral history for them than for any other member of the subfamily. Nonetheless, reconstructing life history characteristics of ancient human is not easy (Opie *et al.* 2012).

Still, this data can be used in an effort to group life history traits as either shared between the human-*Pan* last common ancestor, or derived in the hominin lineage (Robson & Wood 2008). Outgroup analysis is usually a good place to start when trying to determine which traits are shared and which are derived between any two species. This type of analyses yields a “starting point” from which a model can be developed to explain the evolution of derived traits (Lovejoy 2009; Chapais 2013). However the closest living relatives of the human-*Pan* ancestor have mating systems and behaviors that can’t be clearly classified as either *Pan*-like or human-like (Brown 1991). Thus, to explain human/primate life history since the human-*Pan* split, models must be developed that are based on some assumptions (Shultz *et al.* 2011; Opie *et al.* 2012; Plavcan 2012). These assumptions, are in part, based on the use of the fossil record. For example, sexual size dimorphism in ancestral humans yields useful information when inferring past behavior (Plavcan 2000, Reno 2003). This means that the selected starting point of the model determines the ways in which specific fossil traits are associated with the sequential change in life history, from ancestral to modern (Chapais 2013). I will go over some of the



life history traits inferred from the fossil record, in concert with two recent models that attempt to recapitulate the evolution of human social systems.

In *Monogamy, Strongly Bonded Groups, and the Evolution of Human Social Structure*, Bernard Chapais proposes a model based on a single-male polygyny as the ancestral *Pan*-human social system's starting point (Chapais 2013). Chapais first defines the current social system of humans, which he recognizes as a difficult task, since human social structure is concealed by cultural expression (Chapais 2011a). Performing a comparative analysis between human and nonhuman primate societies, Chapais concludes that the current human social system is a federation of multifamily groups socially characterized as monogamous, with strong bonds between groups and lifelong kin recognition (Chapais 2009). He reasons that the *Pan*-human ancestor is most likely gorilla-like in a single-male multi-female polygyny termed a one male unit (OMU), or baboon-like with a grouping of polygyny groups (multi-OMU). In this model, the next step is multi-OMU groups becoming weakly bonded to each other. The third step is a transition towards monogamy, from weakly bonded multi-OMU groupings to weakly bonded multifamily groupings with reduced polygyny. A strengthening of between-group bonds follows. The last step is a transition towards multi-group federations. In summary, Chapais' model has the ancestral human state as polygynous. The transition towards monogamy occurs because of an increased cost of polygyny, and is marked by reduced male-male aggression and increased tolerance of non-familial and subordinate male mating (Reichard & Boesch 2003; Chapais 2011b; Nakahashi & Horiuchi 2012).

C. Owen Lovejoy presents an alternative model in *Reexamining Human Origins in Light of Ardipithecus ramidus*. He proposes the starting point of the ancestral *Pan*-human social system to be multi-male multi-female (Lovejoy 2009). This model includes *Ardipithecus ramidus*, an upright walking non-specialized omnivore with little sexual size dimorphism, as the earliest example of an ancestral human. The model suggests that several factors led towards social monogamy with decreased territoriality, cryptic ovulation, decreased intrasexual agonism and loss of the sectorial canine complex. These factors include upright walking associated with male provisioning, a decrease in both the number of male mating partners and female reproductive rate, and an increase in home range and desire for protein and fat, concealed ovulation, and increased paternal care. The previous changes further allowed for an increase in maternal care and alloparenting, as well as cooperative male patrols with larger males. All of this set the stage for modern human.

#### 1.1.4 Sexual Selection

##### 1.1.4.1 Overview

Natural selection favors the fittest organisms in diverse populations and is the mechanism by which organisms acquire adaptive characteristics. Although asexual reproduction, or reproduction by one parent, produces offspring at a higher rate, sexual reproduction, or reproduction by two parents, persists. In an asexual population, every individual reproduces, but each offspring is a clone and hence any deleterious mutation that occurs will remain through each future generation of that lineage, a process known as

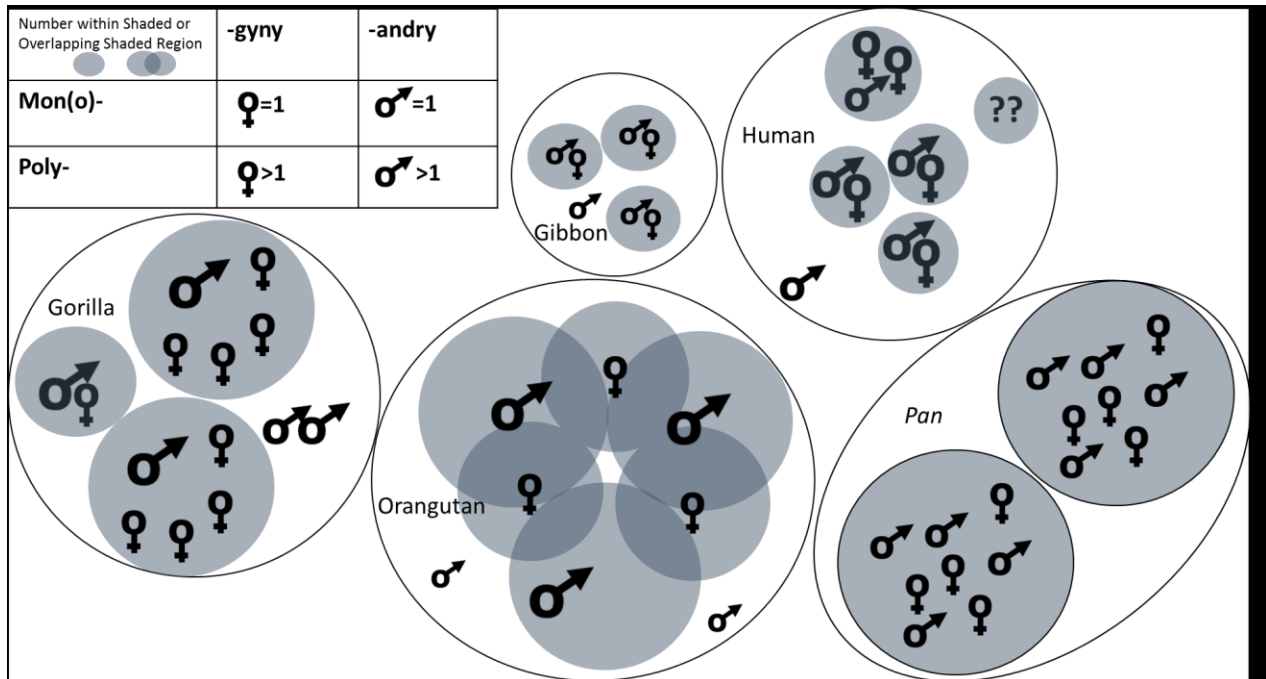
Muller's ratchet (Muller 1950; Felsenstein 1974). Genetic recombination, on the other hand, allows different genetic combinations to exist in every generation. This permits deleterious mutations to be maintained at lower frequencies, and beneficial mutations to be maintained and combined at higher frequencies in any sexually reproducing population (Agrawal 2001). This is the reason why sexual reproduction persists: natural selection acts more efficiently in sexually reproducing populations (Rice & Chippindale 2001). However, natural selection alone does not explain why strikingly different secondary sexual characteristics exist between the sexes in any population (Darwin 1871).

Differences between the sexes that appear to have fitness costs are not explainable by natural selection alone. Charles Darwin postulated that these differences, or secondary sexual characteristics were due to sexual selection caused by competition within (intra-sexual) and between (inter-sexual) the sexes for mating opportunities. This type of selection can be associated with traits that help determine paternity either before copulation (pre-copulatory), or after copulation (post-copulatory selection) (Darwin 1871; Dixson 2012). Sexual selection operates on these characteristics when they lead to greater reproductive success (Andersson 1994) leading to sexually dimorphism (Darwin 1871).

#### 1.1.4.2 Sexual Selection in Hominoids

The mating systems of hominoid species differ in size and structure (Figure 1-5). These varying mating systems are attendant with differences in behavioral and associated morphological features (Table 1-1), which suggest that the great apes have been and are being exposed to different sexually selective forces (Dixson 2012). This hypothesis is

supported through a number of comparative behavioral and physiological analyses, as well as genetic studies.



**Figure 1-5 Mating Systems**

Mating system arrangements for each species, showing the number and sex of individuals within a territory (gray circles) or within overlapping territories. Opportunities for mate acquisition are limited to those within the grey circles, or within overlapping circles. Each system is named after either the number of males or the number of females that can interact within any given territory.

**Table 1-1 Hominoid Behavior and Associated Morphology**

	<u><b>Human</b></u>	<u><b>Pan</b></u>	<u><b>Gorilla</b></u>	<u><b>Orangutan</b></u>	<u><b>Gibbon</b></u>
<b>Mating System</b>	Monogamish: <i>Monogamy with background polyandry and polygyny</i>	Multi-male / Multi-female <i>Polyandry/ Polygyny</i>	Uni-Male/ Multi-female <i>Monandry/ Polygyny</i>	Dispersed <i>Polyandry/ Polygyny</i>	Pair-Bonded <i>Monogamy</i>
<b>Paternity</b>	Known	Unknown	Known	Unknown	Known
<b>Sperm Competition</b>	None-Low	High	None	Moderate	None
<b>Testis to Body Size</b>	Moderate	Large	Small	Moderate	Small
<b>Ejaculate Viscosity</b>	Moderate	Solid	Low	Moderate	Low

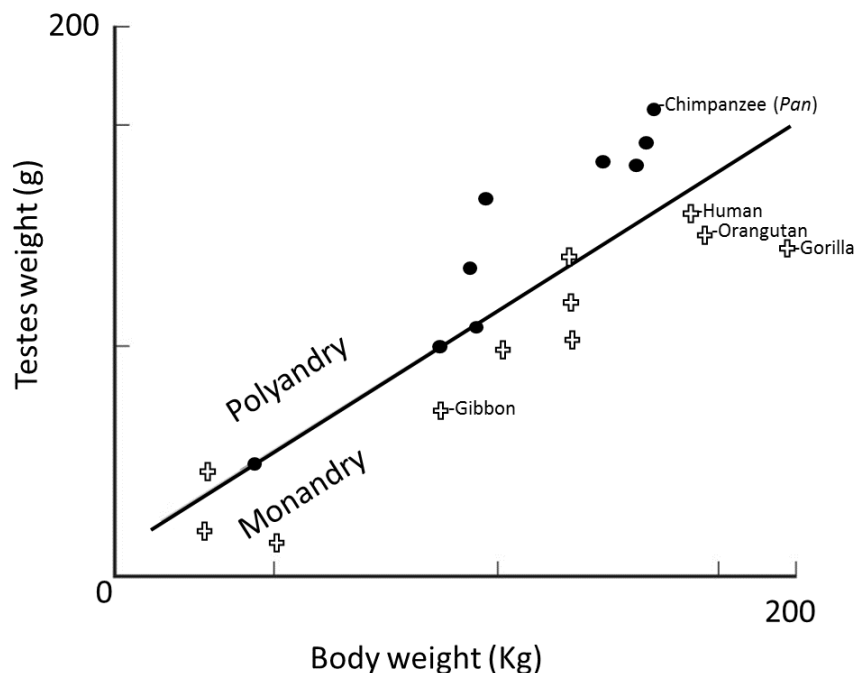
#### 1.1.4.2.1 Intrasexual Selection in Hominoids

Two representative types of intrasexual competition can be described in hominoids: pre-copulatory mate guarding, or post-copulatory sperm competition (Dixson 2012). Mate guarding has been associated with sexual dimorphism in body size. For hominoids, sexual dimorphism in body size and mate guarding are exemplified in gorilla. Across primate species, they occur together to the greatest extent in polygynous monandrous groups, they are moderate in multi-male multi-female groups, and mostly absent in monogamous groupings (Clutton-Brock & Harvey 1977; Clutton-Brock & Harvey 1979). Because the male must ward off other males to ensure paternity, the number of females that he can mate with is limited to those that he can protect from other males and from predation. In these cases, selection favors the larger males. However, body size dimorphism is most likely not a

trait obtained completely through intrasexual selection, as female choice may play a role as well (Clutton-Brock & Harvey 1979; Dixson 2012).

Sperm competition is another type of intrasexual competition associated with specific types of morphology. Chimpanzees and bonobos live in multi-male multi-female groups and have the highest level of sperm competition in hominoids (Dixson 2012). As discussed earlier, females will mate with multiple males, multiple times, during any given period of fertility. In this circumstance, the sperm from multiple males compete for a chance at fertilization, a phenomenon known as sperm competition.

This clearly selective post-copulatory force is associated with traits such as increased testes size, increased seminal volume, increased sperm concentration and motility, seminal plug formation, and decrease or loss in seminal liquefaction (Roussel & Austin 1967; Birkhead & Møller 1998; Dixson & Anderson 2002). For example, large testes compared to body mass is selected for in groups with higher levels of sperm competition (Figure 1-6) (Harcourt *et al.* 1981; Dixson 1995; Harcourt *et al.* 1995). Additionally, semen viscosity increases as sperm competition increase, with plug formation occurring in the seminal plasma of *Pan* (Dixson & Anderson 2002). The plug decreases access to the uterus during subsequent matings in *Pan* (Dixson & Anderson 2002). In human, the viscous ejaculate populates and possibly monopolizes the cervix, allowing sperm to be released over time (Insler *et al.* 1980). Popular media has suggested that this, and other traits like killer sperm, may be selected for in the presence of sperm competition in humans (Ryan 2011); (Baker 1996). However, there is no evidence to suggest that mechanisms like these have been selected for in humans (Moore *et al.* 1999).



**Figure 1-6 Linear regression of testes versus body weight**

A few representative primate species plotted. Block dots represent polyandrous species, white crosses represent monandrous species. Redrawn from Harcourt A.H. et al. (1981) (Harcourt *et al.* 1981)

#### 1.1.4.2.2 Intersexual Selection in Hominoids

Like intrasexual competition, intersexual competition can work at both the pre-copulatory and post-copulatory levels. Pre-copulatory selection with intersexual competition has been traditionally associated with female choice related to adornment, or to secondary sexual characteristics in males (Darwin 1871). In gorilla, females may choose males with healthy genes, or those that protect the harem from predation or infanticide (Caro 2005). This may mean selecting the male with the largest body size, sagittal crest, gluteal muscles, or even just selecting the largest harem (Vanpé *et al.* 2008). The female is

more likely to stick with, or choose, a mate if he is able to show that he can protect her and her offspring (Harcourt & Stewart 2007). Regardless of what type of female choice is working in gorilla populations, it is clear that the largest harems have both the lowest infant mortality and the males with the largest sexually dimorphic features (Breuer 2008).

At the post-copulatory level, intersexual competition is a little more obscure in hominoids than in other clades. In *Drosophila*, for instance, there are multiple examples of this type of copulatory selection, including seminal proteins that reduce female re-mating, or female ovum that are resistant to sperm penetration (Fowler & Partridge 1989; Chapman *et al.* 1995; Holland & Rice 1998). These copulatory traits are clearly beneficial to one sex while having a cost to the other.

In contrast, post-copulatory traits like these are not always obvious in primates. In chimpanzee, females advertise estrous through genital swellings, which induce males to copulate. In bonobos, the females continuously induce sperm competition (Wrangham *et al.* 1996; Dixson 2012). In many primate species, females mediate copulatory behaviors with mating songs, which may extend to post copulatory invitation or to preventing sperm competition (Birkhead & Møller 1998; Maestripieri & Roney 2005). Unlike in other clades, the traits in these examples may be pre-copulatory or post-copulatory depending on when they are expressed.



## 1.1.5 Selection and *Cis*-Regulatory Changes

### 1.1.5.1 Coding versus Noncoding Evolution

In 1975, King and Wilson noted that the anatomical, physiological, behavioral, and ecological differences between human and chimpanzee could not be explained by the biochemical or gene coding differences that existed between these species. They hypothesized that these differences must then be the consequence of gene regulation (King & Wilson 1975). Sequencing technology and early theoretical work, like that of Motoo Kimura's *Neutral Theory of Molecular Evolution* (Kimura 1979), have made it relatively easy to study selection associated with synonymous and non-synonymous mutation within gene coding regions. Consequently, this is where much of the work in understanding adaptive change has focused. It is much easier to understand how mutations affect protein coding or amino acid substitution than it is to even identify mutations that affect gene regulation (Wray *et al.* 2003). Nevertheless, both coding and non-coding changes play an important role in phenotypic evolution. Through the advancement in technology, the exploration into the role of non-coding changes in adaptive evolution has recently become more accessible (Wray 2007).

*Cis*-regulatory regions or elements are part of the DNA sequence of a gene, while factors that bind to *cis*-regulatory regions that referred to as *trans*-acting. Both play a role in gene expression. Mutations causing adaptive changes to *trans*-acting factors can play a role in differential gene expression. Because *trans*-acting factors usually bind to many different genes, mutations to a *trans*-acting factor typically affect multiple processes and phenotypes. A mutation to this type of factor yielding selective advantage to one trait

would likely be linked with other neutral and deleterious changes. This reduces or eliminates the likelihood that selective advantage would favor most changes to trans-acting factors (Raff & Kaufman 1991). On the other hand, *cis*-regulatory changes may target specific aspects of gene function, allowing *cis*-regulatory regions to be more “evolvable” than either coding regions or *trans*-acting factors (Arnone & Davidson 1997). Notably, in closely related species, like the mouse and human, *cis*-regulatory differences overwhelmingly have the largest impact on species differences than any other factor (Coller & Kruglyak 2008; Wilson *et al.* 2008).

#### 1.1.5.2 Testing for Selection in *Cis*-Regulatory Regions

There are some recent examples of *cis*-regulatory or non-coding mutations that have been associated with physiological differences. The difficulty lies in identifying how possible changes to regulatory regions that also affect phenotype occur, so that reasonable models can be established that test for selection within *cis*-regulatory regions (Wray *et al.* 2003). In what follows, I will discuss some of the latest attempts to do this, and highlight some of the relevant methods and findings.

In one example, McLean *et al.* (2011) (McLean *et al.* 2011), examined non-coding conserved deletions, and showed that some of these deletions contained enhancers whose loss led to either anatomical loss or change derived in the human lineage. Using whole genome comparisons, the group identified sites, conserved in mammals and under purifying selection in chimpanzee, which are deleted in humans (hCONDELs). They then provided a descriptive analysis, which included finding an overwhelming majority of

deletions in non-coding regions, and enriched in regions including those involved in steroid hormone receptor signaling and neural function. McLean *et al.* then picked specific candidate hCONDEL based on knowledge of the proximal genes, and proceeded with a functional investigation of the conserved regions within the deleted area. These regions were cloned into a *lacZ* reporter with a basal promoter to test the ability of the region to drive expression in transgenic mice embryos, and noted tissue specific response to the conserved region. This process successfully identified specific enhancer elements, lost during human evolution, and experimentally characterized the connection to tissue specific patterns of expression. Though they acknowledged that there is no way to absolutely determine if the changes were adaptive, this effort makes a strong case for positive selection, connecting specific regulatory deletions to evolutionarily significant phenotypic changes.

In another example, Rockman *et al.* (2005)(Rockman *et al.* 2005) , examined non-coding *cis*-regulatory element using tools developed to study protein coding sequences and *in vitro* reporter assays. The group begins by selecting a gene associated with a uniquely adaptive human trait, which includes a possible upstream regulatory region with polymorphic alleles associated with mental pathologies. They then sequenced this region in multiple human and non-human primates showing that the possible regulatory region has a highly improbable number of substitutions in the human lineage as compared to the other species under a neutral model. Rockman *et al.* concluded that the most likely way that so many new substitutions could become fixed is if positive selection had been acting on the region. Determining the effect of this region, and another repressor element with differences

between species on gene expression required functional evidence. For this, they cloned the human and chimpanzee regions into a luciferase reporter, and to separate the effect of the other element, they also produced chimeric constructs. These constructs were then transfected into a human cell line which normally expressed the gene of interest. The results showed species specific responses, reinforced by the chimeric constructs, connecting the differences found in the regulatory elements to differences at the level of expression. Importantly, although the actual phenotype affected by this gene is still unknown, its association with human evolution, and the use of multiple lines of evidence suggest that the regulatory region has been under positive selection in the human lineage.

Although these two examples illuminate selection in regulatory elements, more examples are needed to determine the genetic and molecular basis for *cis*-regulatory evolution (Wray 2007). In the following chapters, I provide two novel examples of the genetic and molecular basis for *cis*-regulatory evolution between closely related hominid species.

## 2 Evolution of transcriptional regulation of *ACPP* in hominoids

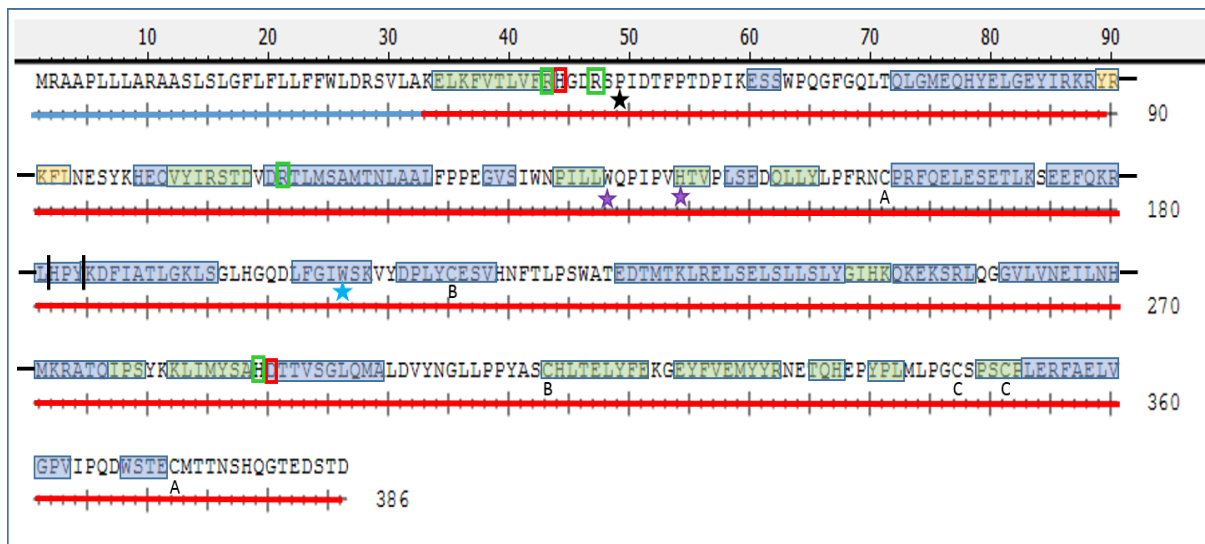
### 2.1 Introduction

#### 2.1.1 ACPP Protein Structure

Prostatic acid phosphatase (ACPP) is an approximately 100kDa homodimer of two non-covalently associated subunits, each of which is approximately 50kDa. Homodimerization is necessary for catalytic activity of the subunits (Kuciel *et al.* 1990). Each subunit has two domains. The larger domain is a seven stranded beta sheet with alpha helices on either side. The smaller domain is composed of six alpha helices (Ortlund *et al.* 2003).

The major ACPP isoforms are the cellular isoform and the secreted isoform. These isoforms have different post transcriptional and post translational modifications which lead to different immunological (Vihko 1979; Lee *et al.* 1984), biochemical (Lad *et al.* 1984), antigenic (Vihko 1979), and glycosylation patterns (White *et al.* 2009).

Secreted ACPP, the form found in seminal plasma, is 354 amino acids in length before after the 32 amino acid signal peptide is cleaved (Figure 2-1). The active residues in the mature protein are His<sup>44</sup> and Asp<sup>290</sup> (Zhang *et al.* 2001). The amino acids upstream of the active residues, Arg<sup>43</sup> and His<sup>291</sup>, as well as amino acids Arg<sup>47</sup> play an active role in substrate binding (Ostanin *et al.* 1994).



**Figure 2-1 ACPP Primary and Secondary Structure**

Reading from left to right, N to C termini, the first 32 amino acids are the signal peptide (blue line), amino acids 33 through 386 are the mature protein (red). The highlighted amino acid secondary structures are beta strands (green), alpha helices (blue), or turns (orange). The amino acids in a box are substrate binding sites (green), or active sites (red). The stars under the amino acids show the amino acids necessary for substrate specificity (black), homodimerization (purple), or structural stability (light blue). The letters A,B,C are located under areas where cysteine pairs form disulfide bonds. The vertical black lines between amino acids separate one alpha helix structure from another (Jakob *et al.* 2000; Muniyan *et al.* 2013).

### 2.1.2 ACPP Activity and Function

ACPP activity and function varies depending on environment in a pH dependent manner. As a member of a family of enzymes known as acid phosphatases, the phosphatase activity of ACPP is optimal in acidic conditions. Orthophosphoric monoesters and phosphorylated proteins are dephosphorylated by this enzyme in environments ranging from pH 3-6 (Zelivianski *et al.* 1998; Brillard-Bourdet *et al.* 2002). In addition to the phosphatase activity, ACPP acts as a protease, hydrolyzing the cleavage of semenogelin I

derived substrates (Brillard-Bourdet *et al.* 2002). In contrast to the phosphatase activity, this proteolytic activity occurs in basic conditions with optimal activity around a pH of 9 (Brillard-Bourdet *et al.* 2002).

The acid phosphatase activity of ACPD functions on multiple targets. It is a tyrosine phosphatase that also has nonspecific phosphatase activity catalyzing the dephosphorylation of multiple seminal substrates including AMP, lysophosphatide, and ErbB-2 (Chuang *et al.* 2010). The activity on AMP is in association with extracellular, or seminal, 5'-nucleotidase activity (Zylka *et al.* 2008). This makes ACPD active in the adenosine metabolic process. The production of adenosine stimulates a pathway which has anti-nociceptive or chronic pain relieving effects (Zylka *et al.* 2008). The lysophosphatidic acid phosphatase activity is extracellular, or seminal, and is responsible for dephosphorylating and deactivating lysophosphatidic acid, a lipid mediator. Importantly, the lysophosphatidic acid is deactivated by ACPD, and may play a role in immune function, fertility, and uterine egg implantation (Tanaka *et al.* 2004). The phosphatase activity on cellular ErbB-2 targets the tyrosine 1221/2, and decreases androgen independent prostate cell proliferation, with reduced cellular ACPD associated with increased prostate cell proliferation (Chuang *et al.* 2010).

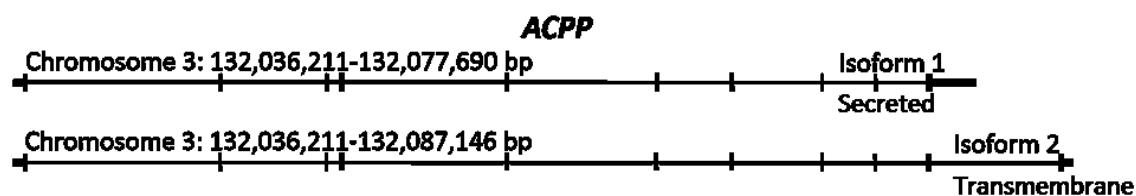
The proteolytic role of ACPD is associated with its extracellular roles associated with seminal liquefaction (Brillard-Bourdet *et al.* 2002). Although many substrates have been found, a clear physiological function has yet to be determined (Kong & Byun 2013). The proteolytic activity preferentially cleaves at Tyr136, Tyr292 and Gln266 of semenogelin I.

This cleavage occurs at a neutral to slightly basic pH rather than the acidic pH at which phosphatase activity is greatest (Brillard-Bourdet *et al.* 2002).

Although ACPD concentration is proportional to sperm motility in seminal fluid, there is an inversely proportional relationship between the seminal concentrations of sperm and ACPD in humans (Dave & Rindani 1988; Singh *et al.* 1996). The mechanism for this correlation is not known, but ACPD level is an effective determinant of fertility in human males (Singh *et al.* 1996).

### 2.1.3 ACPD Gene

The ACPD gene is found in humans on the q arm of chromosome 3. Two isoforms are produced from this single gene. The smaller of the two, isoform 1, has 10 exons and encodes the secreted or seminal form of ACPD. The second and longer of the two, isoform 2, with 11 exons, is alternatively spliced and contains a transmembrane domain (Figure 2-2) (Winqvist, Virkkunen *et al.* 1989, Li and Sharief 1993).



**Figure 2-2 Human ACPD Gene Structure**



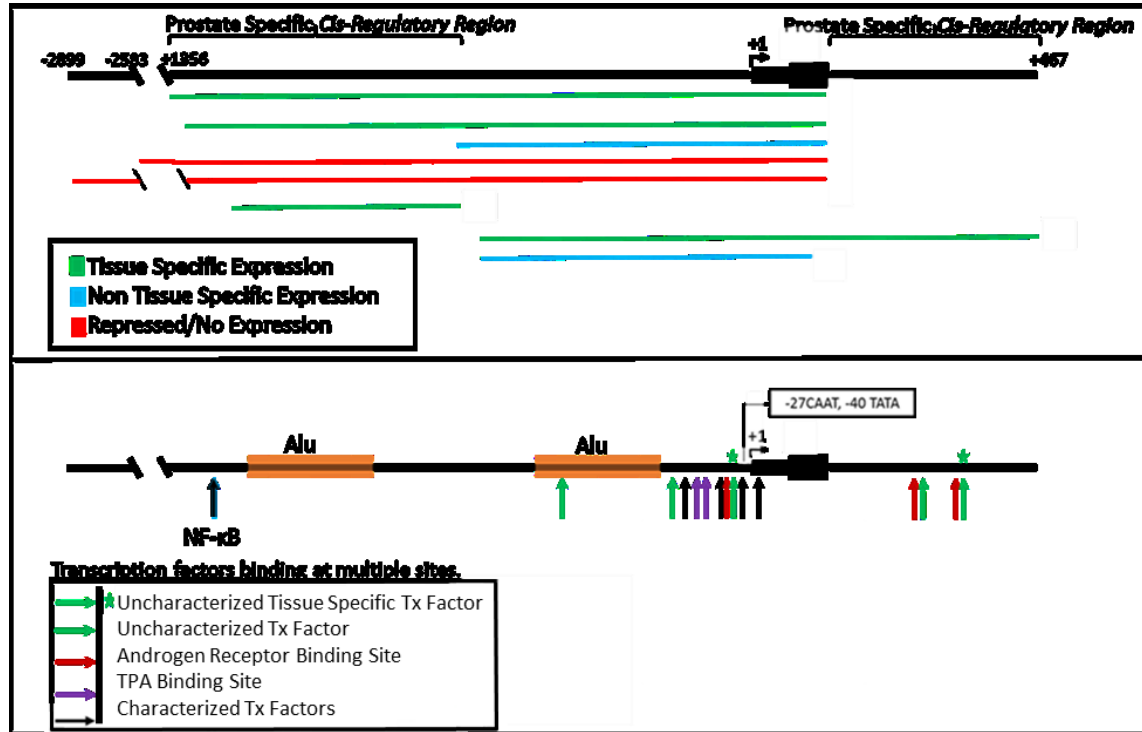
#### 2.1.4 ACPD Regulation

The regulatory elements of any eukaryotic gene can extend thousands of base pairs upstream, downstream, or within the gene itself. Two putative, prostate specific promoter regions were identified within -1356 to +87 (Zelivianski *et al.* 1998), and also within -734 to +467bp (Shan *et al.* 1997) where +1 is the start of transcription. Of the two ACPD regulatory regions tested a core promoter was identified between -779 and +87 (Zelivianski *et al.* 2000), and between -734 and +50 (Shan *et al.* 2003). Regions that down-regulate transcription in prostate cell lines extend from -2899 to -2583, -2583 to -1305 and -1668 to -1356(Zelivianski *et al.* 2000; Zelivianski *et al.* 2002). An enhancer that is prostate specific in the prostate cell lines PC-3 and DU-145, and in a transiently transfected mouse model, extends from -1258 to -779 (Zelivianski *et al.* 2002). Using mouse and prostate cancer cell lines, another region, -734 to +467, was shown to have prostate specific expression over that of the core promoter, -734 to +50 (Shan *et al.* 2003). This makes a prostate specific enhancer element or continuous transcription factor binding likely to occur between +50 and +467 (Shan *et al.* 2003) (Figure 2-3, Top).

ACPD is expressed in the prostate but can also be found in the bladder, kidney, pancreas, lung, cervix, testis, and ovary. Weak expression has also been detected in other tissues, although these transcripts are at least one to two orders of magnitude less than what is expressed in the prostate (Graddis *et al.* 2011).

Although specific trans-acting factors and promoter landscape features associated with ACPD expression are not yet fully understood, many have been identified within the putative prostate specific promoter region, -1356/+467 (Figure 2-3, Top Brackets). In this

region there are two Alu repetitive elements upstream of the transcription start site (Sharief & Li 1994). Upstream of both repetitive elements NF- $\kappa$ B has been shown to bind the hexanucleotide, AGGTGT (Zelivianski *et al.* 2004). This is the first time that NF- $\kappa$ B has been shown to bind to this sequence (Zelivianski *et al.* 2004). There are three androgen response elements (AREs) in the region driving both up-regulation and down-regulation of ACP expression (Banas *et al.* 1994; Shan *et al.* 2003). AREs are DNA sequences that bind the androgen receptor, a nuclear hormone receptor that translocates to the nucleus after binding androgen. There are five regions in the promoter that contain an element with the GAAAATATGATA sequence which is associated with androgen dependent transcription. Two of these elements have been shown to have prostate specific activity, containing a weak association with the AR-USF2 complex (Shan *et al.* 2003; Shan *et al.* 2005) (Figure 2-3, Bottom).



**Figure 2-3 ACPP Regulation**

TOP: Putative cis-regulatory regions tested in prostate tissue (LNCaP, PC-3, DU 145) and non-prostate tissue (HeLa, WI-38, A431, T47D, A-549) cell lines. The colored lines represent the portions of the promoter region tested with reporter constructs in prostate tissue, and non-prostate tissue cell lines. Regions upstream and downstream of transcription start site have been identified as prostate specific cis-regulatory regions. Bottom: Transcription factors and their binding sites with repetitive elements highlighted in orange. Green arrows indicate an uncharacterized nuclear factor with GAAAATATGATA-Like binding affinity. Figure combines the work of: (Zelivianski *et al.* 1998) (Zelivianski *et al.* 2000) (Zelivianski *et al.* 2002) (Shan *et al.* 1997; Shan *et al.* 2003)

The UCSC human genome assembly and its custom tracks provide a wide variety of information pertaining to *ACPP* regulatory region (Kent *et al.* 2002). The ENCODE DNaseI hypersensitivity Clusters from 125 cell types denotes an open or DNaseI sensitive site from about -200 to +650 of the transcription start site with the transcription factor ChIP-seq

from ENCODE indicating possible transcription factor binding sites from -700 to +800 of the transcription start site (Rosenbloom *et al.* 2013). Three repetitive elements are indicated, two Alu's from -1149 to -815 and from -530 through -215, and one MIR from -186 to -97 (Jurka 2000).

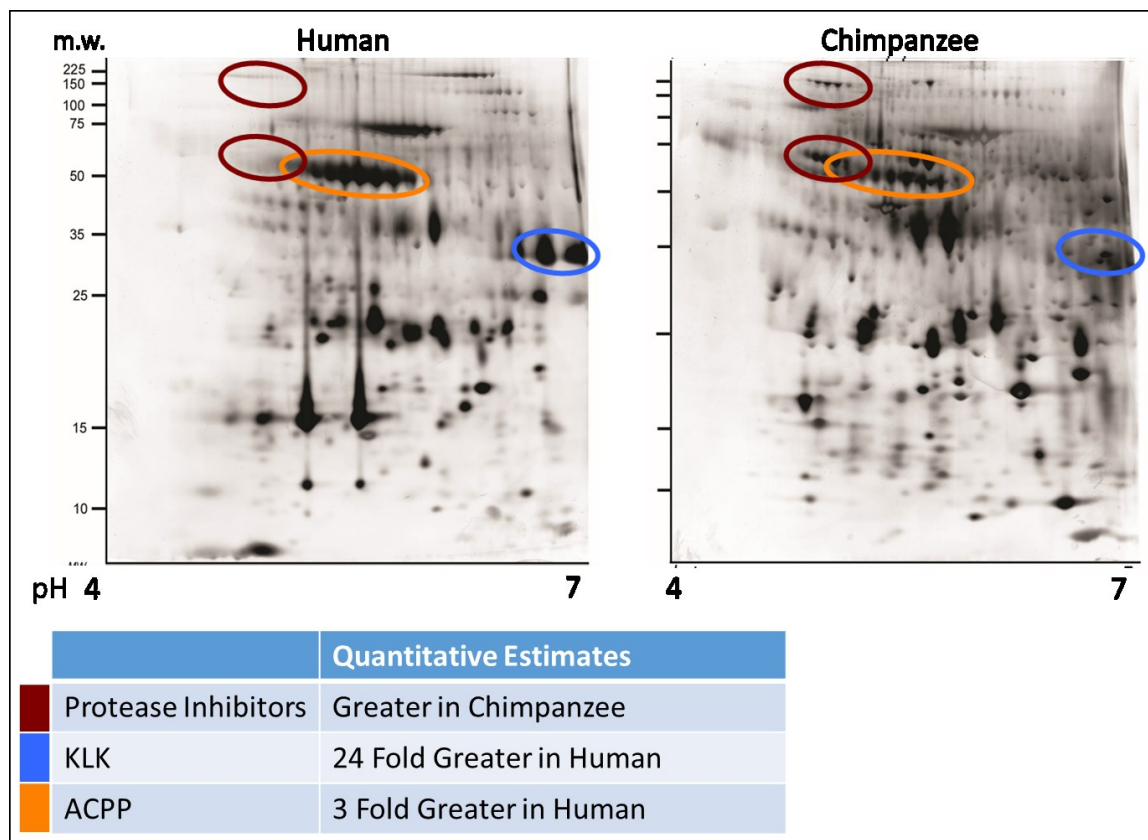
### 2.1.5 Differences between Species

There is a high degree of variability in the mating behaviors of different hominoid species and their associated anatomical correlates. Genes associated with these behaviors and anatomical adaptations are predicted to show signs of positive selection in species with increased sperm competition (Wong 2010). Signs of positive selection have been found in the reproductive genes of hominoids (Jensen-Seaman & Li 2003; Dorus *et al.* 2004; Clark & Swanson 2005; Carnahan & Jensen-Seaman 2008), but as a whole the hominoid seminal protein coding regions do not show an increase in rates of amino acid substitutions when compared to coding regions of non-reproductive genes (Carnahan-Craig & Jensen-Seaman 2014).

King and Wilson (1975) noted that the vast behavioral and anatomical differences between chimpanzee and human could not be accounted for by the small degree of amino acid sequence divergence found between the two species. They hypothesized that between two closely related species, it is more likely that these differences stem from changes in the mechanisms controlling gene expression rather than changes in the amino acid composition of any given protein (King & Wilson 1975).

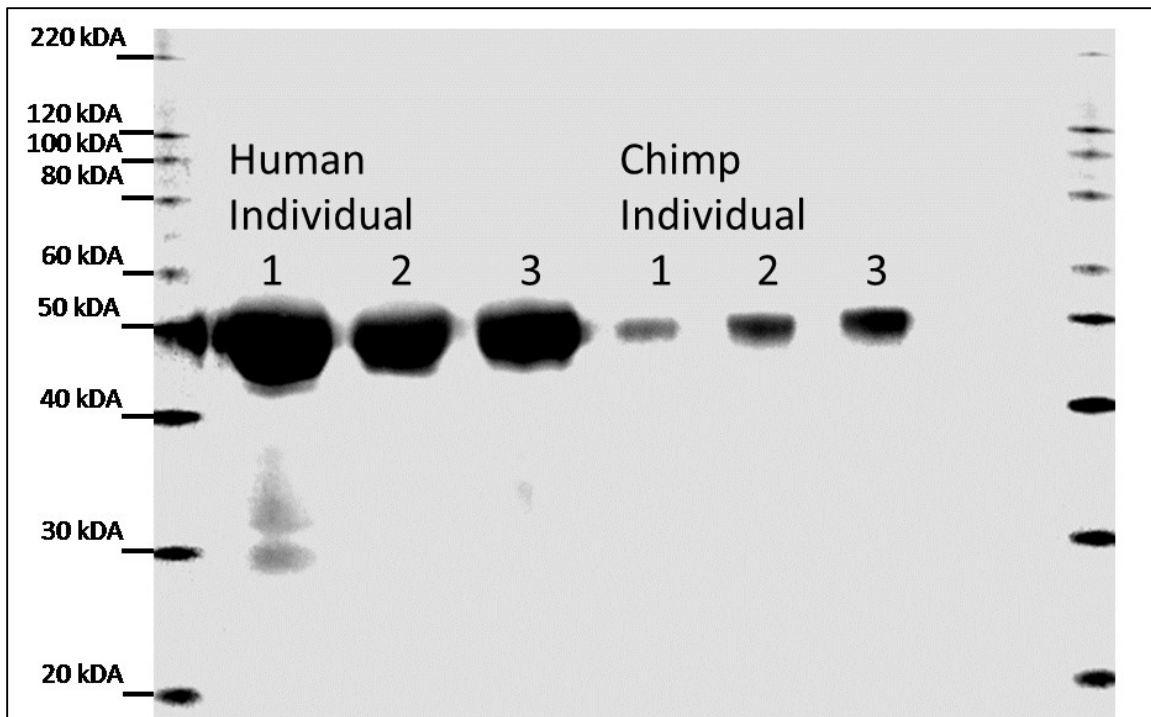
Proteins associated with seminal dissolution are differentially expressed between human and chimpanzee (Figure 2-4) (Chovanec & Jensen-Seaman, unpublished data).

Notably, ACPP is found at much higher concentrations in human seminal plasma (Figure 2-5) (Colvin & Jensen-Seaman, unpublished data). If the current physiological differences are associated with gene regulation, then they must have occurred in one or both of the species since the split from their common ancestor.



**Figure 2-4 Differential concentrations of seminal derived proteolytic molecules.**

Shotgun liquid chromatography/tandem mass spectroscopy and 2D Gels of seminal plasma identify and yield quantitative estimates of plasma derived proteins (Chovanec & Jensen-Seaman, unpublished data).



**Figure 2-5 Western blot of ACPD derived from seminal plasma.**

Western blot from three human samples and three chimpanzee samples, all derived from separate individuals (Colvin & Jensen-Seaman, unpublished data).

Sequence comparison makes it relatively straightforward to understand how coding mutations that affect protein sequence, post-transcriptional processing, and post-translational processing lead to phenotypic variation between orthologous genes. This had led to a coding region bias in work based on understanding this variation. It is still relatively difficult to use sequence comparison alone to understand how mutations affect gene regulation. Cis-regulatory mutations affect transcription and post-transcriptional processing (Wray 2007). It still remains unclear if and how these differences play a role in differential transcription of *ACPD* between species and importantly, at which point, or in which species, these differences arose. Regardless, the most likely change to occur between

closely related species that would cause such a difference in relative abundance would be the cis-regulatory (King & Wilson 1975; Wray *et al.* 2003).

## 2.2 Materials and Methods

### 2.2.1 PCR

#### 2.2.1.1 Genomic Amplification

PCR amplification was performed in 20 $\mu$ l reactions containing 2 $\mu$ l of 10x Taq Buffer “Advanced” with 15mM Mg<sup>++</sup> (Eppendorf), 200 $\mu$ M dNTPs, 1 $\mu$ M forward and reverse primers, about 5ng of genomic template, 0.5 units of Taq polymerase, and molecular biology grade water to 20 $\mu$ l. DMSO or extra Mg<sup>++</sup> were added as needed. The amplification reaction included an initial 2min. melt at 94°C followed by 35 cycles of 94°C for 30 seconds, 55°C for 1minute, and 72°C 30sec/kb. After cycling, the reaction was finished at 72°C for 10 minutes, and 4°C storage.

High fidelity amplification was performed to reduce the likelihood of errors being introduced. The amplification was performed in a 20 $\mu$ l reactions containing 2 $\mu$ l of 10x iTaq™ buffer (Bio-Rad) with 40 $\mu$ M Mg<sup>++</sup>, 200 $\mu$ M dNTPs, 1 $\mu$ M forward and reverse primers, about 5ng of template, 1 unit of iTaq™ DNA polymerase (Bio-Rad), and molecular biology grade water to 20 $\mu$ l. DMSO or extra Mg<sup>++</sup> were added as needed. The amplification reaction included an initial 30 second melt at 95°C followed by 35 cycles of 98°C for 10 seconds, and 56° 15 seconds, 72°C 30sec/kb. After cycling, the reaction was finished at 72°C for 10 minutes, and 4°C storage.



### 2.2.1.2 Colony PCR

The amplification was performed in 20 $\mu$ l reactions containing 2 $\mu$ l of 10x Taq Buffer Advanced with 15mM Mg<sup>++</sup> (Eppendorf), 200 $\mu$ M dNTPs, 1 $\mu$ M forward and reverse primers, 0.5 units of Taq polymerase, and molecular biology grade water to 20 $\mu$ l. DMSO or extra Mg<sup>++</sup> were added as needed. The colonies were picked with a sterile toothpick, dipped into 10 $\mu$ l of reaction mix without polymerase, and then streaked onto a replicate plate. The reaction was then brought to 98°C for five minutes to lyse cells, then returned to ice. The remaining 10 $\mu$ l of reaction mix with polymerase was added to the reaction. The amplification reaction included an initial 2 minute melt at 95°C followed by 35 cycles of 98°C for 10 seconds, and 56° 15 seconds, 72°C 30sec/kb. After cycling, the reaction was finished at 72°C for 10 minutes, and 4°C storage.

### 2.2.2 Sequencing

All sequencing reactions were run at 20 $\mu$ l using 1 $\mu$ l of BigDye® Terminator v3.1, 4 $\mu$ l of sequencing buffer, ~50ng purified product, 3.2pmol of primer, and brought to volume with molecular biology grade water. The reaction included 35 cycles of 96°C for 10 seconds, and 50° 5 seconds 60°C 4 minutes. After cycling, the reaction was finished at 68°C for 10 minutes, and 4°C storage.

Sequencing reactions were purified over packed sephadex slurry columns. The columns were packed by adding 550 $\mu$ l of sephadex slurry into well of a 96 well column plate and spun for 3 minutes at 850 x g. The samples were then loaded onto the packed sephadex, and into a 96 well plate by spinning at 850 x g for 4 minutes. The samples were then heated

at 98°C for 2 minutes and then cold shocked at 4°C for 2 minutes before analyzing the samples on an Applied BioSystems Avant3130.

## 2.2.3 Construct Design

### 2.2.3.1 Obtaining Region of Interest

The candidate regions were obtained using high fidelity amplification from the genomic DNA of human (*Homo sapiens*), chimpanzee (*Pan troglodytes*: PR496), bonobo (*Pan paniscus*: PR251), gorilla (*Gorilla gorilla*: Chipua), and orangutan (*Pongo pygmaeus*: WGA). Primers (Table 5-1) were designed from the human genomic region spanning -1309 to +350bp from the *ACPP* transcription start site, and incorporate *Acc65I* and *HindIII* restriction enzyme sites to allow cloning into the pGL4.10 reporter vector (

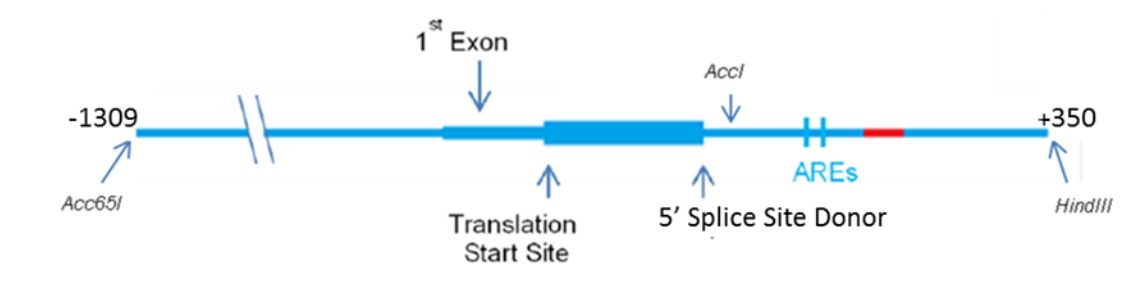


Figure 2-6).

### 2.2.3.2 TOPO® TA Cloning and Screening

The amplified product was gel purified on a crystal violet agarose gel (1%) and column purified using the Wizard® SV Gel and PCR Clean-Up System (Promega). A reaction was set

up consisting of 10µl or ~20ng of the purified product, 2µl of 10x Taq Buffer Advanced with 15mM Mg<sup>++</sup> (Eppendorf), 200µM dNTPs, 1µM forward and reverse M13 primers, 0.5 units of Taq polymerase, and molecular biology grade water to 20µl. The reaction was incubated at 72°C for 10 minutes to add A's to the 3' ends of the PCR product, and then it was placed on ice.

The product (4 µl) was added to a reaction containing 1µl of salt solution (1.2M NaCl 0.06M MgCl<sub>2</sub>) and 1µL of TOPO® vector. The reaction was then incubated at room temperature overnight. The reaction was transformed into One Shot® competent cells by adding 2µl to one vial of cells, then incubated for 30 minutes on ice. Cells were heat shocked for 30 seconds at 42°C then returned to ice. Next, 250µl of S.O.C. media was added to cells, and incubated in a 37°C shaker/incubator for 1 hour. The transformed cells (25µl, 75µl, and 200µl) were then spread onto kanamycin-LB agar plates, and incubated overnight at 37°C. Blue-white screening was used to select for colonies positive for the TOPO-ligated insert. Positive (white) colonies were plated on a kanamycin-LB replicate plate and then colony PCR with M13 primers was used to produce amplified product of the ligated products. Sequencing was performed on amplified product of similar size to that of the amplified insert to insure that colonies containing the recombinant insert of interest have been manufactured.

### 2.2.3.3 Digest

Colonies containing the insert cloned into a TOPO vector, and colonies containing pGL4.10, were grown overnight in 3ml kanamycin- LB (55µg/mL) in a 37°C

shaker/incubator. Plasmids were purified from the overnight incubations using the Qiagen QIAprep Spin Miniprep Kit. For each plasmid preparation, two single digests and one double digest were performed. Each 20µl digest reaction contains 10µl of plasmid DNA, 2µl of 10x digest buffer, and either 1µl of *HindIII*, or 1µl *Acc65I*, or 1µl of both *HindIII* and *Acc65I*. The reactions were then placed at 37°C for 3 hours. The digest containing pGL4.10 was then 5' dephosphorylated with two additions of 5µl CIAP in CIAP reaction buffer for two 30 minute periods at 37°C. The products were then run on a 1% agarose gel stained with ethidium bromide to verify that the digest worked. They were then purified by running the remaining product over another 1% agarose gel stained with crystal violet. The band of interest was then removed and processed using the Wizard® SV Gel and PCR Clean-Up System (Promega).

#### 2.2.3.4 Ligation, Transformation and Screening

The digested product and pGL4.10 vector were ligated at a 3:1 ratio in a 20µl reaction containing 1x ligation buffer and 1µl of ligase. This reaction was run overnight at 16°C. The reaction was then transformed into TG4 *E. coli* competent cells, by adding 2µl of reaction to one vial of cells, and incubated for 30 minutes on ice. The cells were then heat shocked for 120 seconds at 42°C and returned to ice. Next, 500µl of LB solution was then added to the cells. Cells were incubated shaking at 37°C for 1 hour. The transformed cells (25µl, 150µl, and 250µl) were then spread onto ampicillin-LB agar plates, and incubated overnight at 37°C. Colony PCR and sequence screening with pGL4.10 vector primers were used to ensure that the correct insert was ligated into the vector in the correct direction, as

described above. After selecting the desired colonies, a 1ml aliquot of culture was grown up at 37°C overnight, then cryopreserved in a 30% glycerol solution at -80°C

#### 2.2.3.5 Mutagenesis

In order to modify the ACP insert cloned into pGL4.10 reporter vectors, the Agilent Technologies QuikChange II XL Site-Directed Mutagenesis Kit was used along with primers designed to knock out a translation start site (ATG to AAG), and to knock out a 3' splice site (TGGT to TCCT)(table). Each mutagenesis reaction uses 10ng of template and 15pmol of each primer, along with the supplied buffer and dNTPs. The reaction was brought to 49µl total volume with molecular biology grade water. Just before beginning the amplification reaction 1µl of *Pfu* Ultra High Fidelity DNA polymerase was added. The amplification reaction included an initial 30 second melt at 95°C followed by 14 cycles of 95°C for 30 seconds, and 55° 1 min., 68°C 8 minutes. After cycling, the reaction was finished at 68°C for 10 minutes, and 4°C storage. When amplification reaction ends, 1µl *DpnI* was added and kept at 37°C for 1 hour. The cells were then transformed by pipetting 2µl of the reaction into XL10-Gold Ultracompetent cells supplied with the kit. The cells were incubated for 30 minutes on ice, then they were heat shocked for 30 seconds, and then incubated on ice for 2 minutes before adding 500µl NZY+ broth. The cells were placed at 37°C for 1 hour before plating on LB-ampicillin plates, and incubated overnight at 37°C. The colonies were then picked, PCR screened, and sequence verified.

## 2.2.4 Reporter Assays

### 2.2.4.1 LNCaP Cell Line

The cell line used for all reporter assays was the LNCaP clone FGC (ATCC® CRL-1740™) cell line from ATCC®. Cell lines arrive frozen on dry ice from ATCC® and were placed in liquid nitrogen storage for future thawing and recovery, or were immediately thawed and recovered upon receipt.

### 2.2.4.2 Thawing and Recovery of Cells

Cryopreserved cells were removed from storage, and placed into a 37°C water bath for one minute. Thawed cells were then resuspended in a 15ml conical tube containing 2ml of RPMI-1640 media prepared with 10% fetal bovine serum (FBS), and 100 I.U./mL of both penicillin and streptomycin (whole media), and centrifuged for 10 minutes at 200 x g. The supernatant was removed, and the pellet was again resuspended in 5ml whole media and placed into a T-25 Poly-D-Lysine coated culture flask, in a 37°C, 5% CO<sub>2</sub> incubator.

### 2.2.4.3 Trypsinizing and Subculture

The cells were removed from the incubator. The old medium was removed with a sterile serological pipette. The monolayer was washed by adding then removing 1 ml of 37°C Hank's buffered salt solution (HBSS). Warmed Trypsin-EDTA (1ml) was then added, and the cells were placed back in the incubator for about 5 minutes, or until the cells were no longer adherent to the monolayer. Trypsin activity was then stopped by adding 2ml of

whole media. The cells were then brought into suspension by pipetting vigorously. This helps removes the remaining cells that were either adherent to each other or to the bottom of the flask. After the cells were in suspension, 1ml was added to each of four new culture flasks, and then 4ml of whole media was added to each. The culture flasks were then returned to the 37°C, 5% CO<sub>2</sub> incubator.

#### 2.2.4.4 Preparation of 12 well plates for transient transfection.

Cells were trypsinized as above. Once the cells were in suspension, and the trypsin activity had been stopped by the addition of 2ml of whole media, the cell density was determined. Cells were prepared for counting by adding 200µl of cell suspension to a tube containing 300µl trypan blue, and 500µl of HBSS. This creates the 1:5 cell preparation dilution. From this, 50µl was pipetted onto either side of a hemocytometer. Viable cells were then counted from 10 of the 0.16mm x 0.16mm boxes on the reading field, 5 boxes per side of the hemocytometer. The average number of cells per 0.16mm x 0.16mm square was then calculated by dividing the total number of cells counted by 10. The average number of cells per milliliter was then calculated, first by adjusting for the 1:5 dilution factor by multiplying by 5 and then by correcting for volume (each square in hemocytometer was 0.1µl) by multiplying by 10,000. This equals the number of cells per milliliter. Next, 200,000 cells were added to each well of a 12 well corning plate, after which each well was brought to 1ml final volume using whole media. The plates were then returned to the 37°C, 5% CO<sub>2</sub> incubator.

#### 2.2.4.5 Transient Transfection

All transient transfections were performed 24 hours after plating cells. Each experiment contains a transfected experimental construct (a regulatory region in a pGL4.10 reporter plasmid), a transfected empty construct (a circularized pGL4.10 reporter plasmid without a regulatory region), and a mock transfection containing no DNA. For the experimental constructs, 1µg of DNA and 3µl of FuGENE® HD transfection reagent were added to RPMI media (without FBS) for a total volume of 50µl, and then incubated at room temperature for 5 minutes before being added to one well of the plated cells. For the empty construct, the molar equivalent of the test construct and 3µl of FuGENE® HD transfection reagent were added to RPMI media for a total volume of 50µl, and then incubated at room temperature for 5 minutes before being added to one well of the plated cells. For the mock transfection, 3 µl FuGENE® HD transfection reagent was added to RPMI media for a total volume of 50µl, then incubated at room temperature for 5 minutes before being added to one well of the plated cells. For the experimental and empty constructs, the transfection mix was scaled up and added to three triplicate wells. The transfected cells were then returned to the 37°C CO<sub>2</sub> incubator for 48 hours.

If the experiment includes induction by the synthetic androgen R1881, the transfected cells were removed from the incubator after 24 hours, and either R1881 was added to a final concentration of 10nM or the equivalent volume of the vehicle control (EtOH). Cells were returned to the 37°C CO<sub>2</sub> incubator for 24 hours.



#### 2.2.4.6 Luciferase Activity Quantification

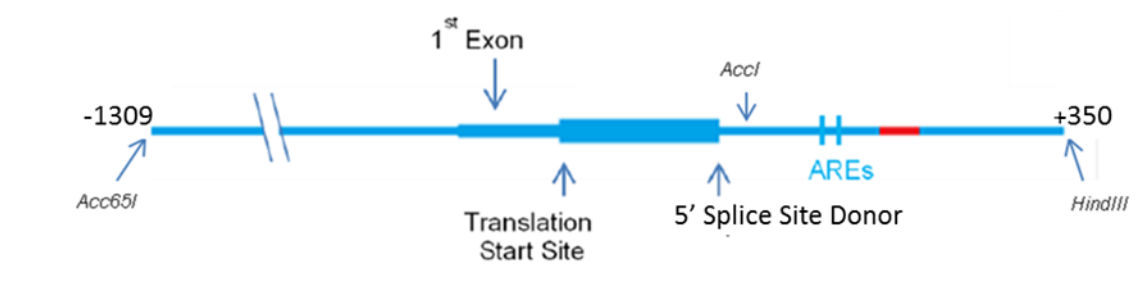
Cells were lysed 48 hours after transfection. To do this, the media was removed from each well and washed by adding then removing 250µl of PBS. Then 250µl of passive lysis buffer was added (Promega Luciferase Assay System), and the plate was placed on a room temperature shaker for 15 minutes. The cell lysate was collected from each well for immediate measurement of activity, or saved at -20°C for later analysis.

From the cell lysates, 20µl was added to a 96 well plate. The plate was placed into the Veritas™ 96 well Microplate Luminometer. The luminometer adds 100µl of luciferase to a well. After a 2 second delay, the luminescence was read for a 10 second period. This process was repeated for each well. Luciferase activity was measured in relative light units. For all data, background luminescence was accounted for by subtracting the average relative values produced from the wells containing the lysate from the mock transfection from the relative value of the wells containing the lysate of the test constructs or the promoterless vector. After adjusting for background luminescence, the adjusted experimental construct values were normalized by dividing by the adjusted promoterless vector values.

#### 2.2.5 Experimental vectors used.

The following putative promoter inserts were placed into reporter vectors, and sequence verified, by aligning to the genomic sequences from which each insert originated. The base promoter sequences were available as an alignment (Figure 5-1 ACPP Putative Promoter Alignments. All insert variants (Figure 2-7) with their names and specific

modifications (Table 2-1) were placed into the multiple cloning site of the pGL4.10 vector between the synthetic poly(A) signal and the reporter coding region (*luc2*) with the same 5' to 3' orientation to *luc2* as exists *in vivo* to ACP. The insert archetype from which all other inserts were derived is shown in Figure 2-6.



**Figure 2-6 Basic Design of Putative ACP Regulatory Region Insert**

#### 2.2.5.1 ACP Reporter

All human, chimpanzee, bonobo, gorilla, and orangutan ACP regulatory regions were amplified from -1309bp upstream to +350bp downstream of the transcription start site using genomic DNA, and primers designed to contain an *Acc65I*, or a *HindIII* restriction site. These inserts were then ligated into the pGL4.10 reporter vector (Figure 2-7 i.).

#### 2.2.5.2 ΔΔACP Reporter

All human, chimpanzee, bonobo, gorilla, and orangutan ACP reporters had their translation start site and 5' splice site donor knocked out through mutagenesis as previously described. The ACP regulatory regions were then amplified from -1309bp upstream to +350bp downstream of the transcription start site using genomic DNA using

the primers designed to contain an *Acc65I*, or a *HindIII* restriction site. These inserts were then ligated into the pGL4.10 reporter vector (Figure 2-7 *ii.*).

#### 2.2.5.3 Chimeric Human and Chimp ACP<sup>P</sup> Reporter

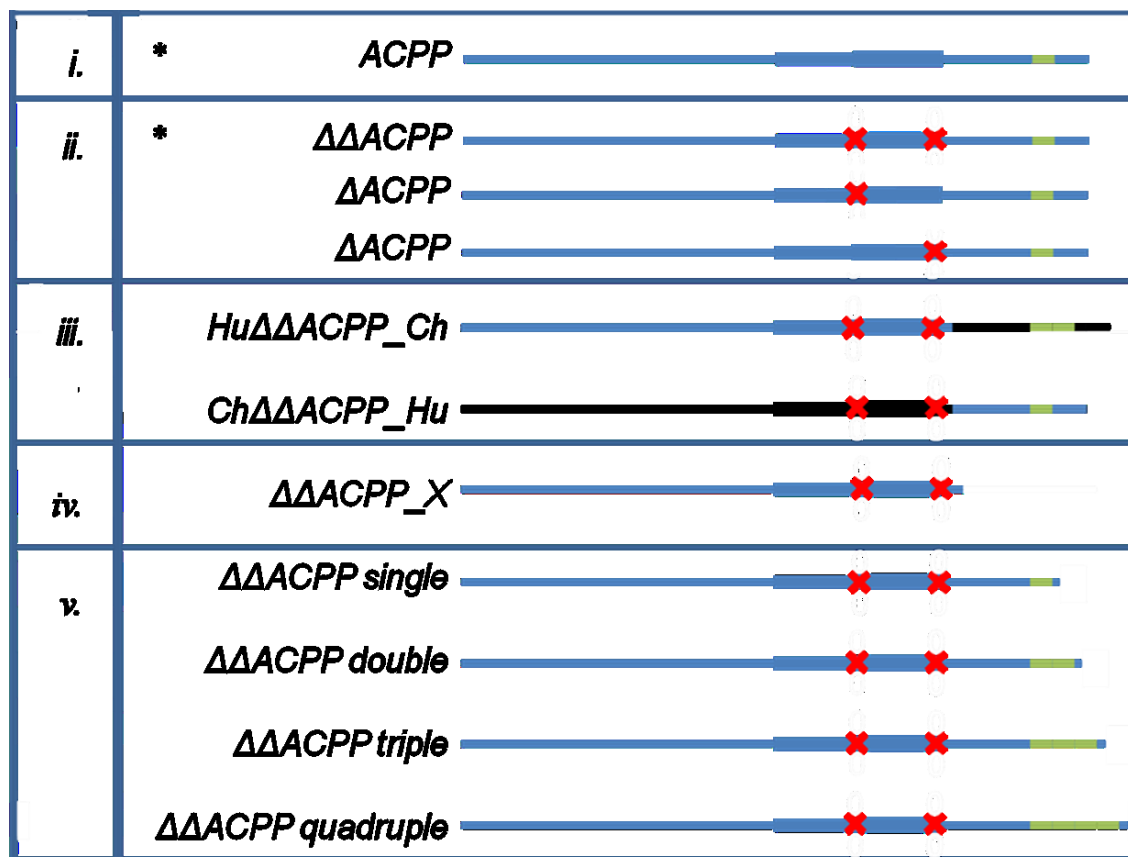
The human ACP<sup>P</sup> and the human  $\Delta\Delta$ ACP<sup>P</sup> were chimpanized, and the chimpanzee ACP<sup>P</sup> and  $\Delta\Delta$ ACP<sup>P</sup> were humanized by swapping the 3' regions of each construct. This was done by digesting the constructs at the ACP<sup>P</sup> endogenous *AccI* recognition site 4bp upstream from the 3' splice donor, and at the *HindIII* site engineered into the 3' terminus of the insert. After digestion, the plasmids and digested fragments were purified, and the inserts were ligated into the opposing species reporter constructs (Figure 2-7 *iii.*).

#### 2.2.5.4 $\Delta\Delta$ ACP<sup>P</sup> Truncated Reporter

The human and chimpanzee  $\Delta\Delta$ ACP<sup>P</sup> constructs were truncated by digesting the region between the endogenous *AccI* recognition site 4bp upstream from the 5' splice donor, and at the *HindIII* site, as above. The vector was then purified from digest product using a 1% agarose crystal violet gel, and the Promega Wizard SV Clean Up System. The vector ends were blunted by setting up a high fidelity amplification with iTaq as described, without the addition of primers, and only placing the reaction at 68°C for 10 minutes. The vector ends were then ligated (Figure 2-7 *iv.*).

#### 2.2.5.5 Human $\Delta\Delta ACP$ P Single, Double, Triple, and Quadruple Repeat.

Chimpanzees and bonobos contain a tandem duplication of 16bp in the first intron, within the putative promoter cloned into the above constructs (see Results). The human  $\Delta\Delta ACP$ P constructs with varying numbers of this 16bp region were produced using a PCR reaction with primers designed to include the specific number of 16bp regions, and using the Human  $\Delta\Delta ACP$ P reporter as template. Each clone was named Human  $\Delta\Delta ACP$ P (single, double, triple, or quadruple) for the number of times the region of interest appears within the insert (Figure 2-7 v.).



**Figure 2-7 *ACPP* Construct Inserts**

Reporter inserts are represented above. The Constructs with a \* are made for five different species, including human, chimpanzee, bonobo, gorilla, and orangutan. The red X represents a knocked out translation start site, or a knocked out 5' splice site donor.

**Table 2-1 Clone Names**

	Construct Names	Up - stream	Down - stream (TSS)	Total Size (Bp)*	Modifications
A	HuACPP	-1309	+350	1659	None
B	ChACPP	-1309	+350	1675	None
C	BoACPP	-1309	+350	1675	None
D	GoACPP	-1309	+350	1659	None
E	OrACPP	-1309	+350	1659	None
F	HuACPP ΔATG	-1309	+350	1659	*
G	HuACPP ΔSp1	-1309	+350	1659	**
H	HuΔΔACPP	-1309	+350	1659	*/**
I	ChΔΔACPP	-1309	+350	1675	*/**
J	BoΔΔACPP	-1309	+350	1675	*/**
K	GoΔΔACPP	-1309	+350	1659	*/**
L	OrΔΔACPP	-1309	+350	1659	*/**
M	HuACPP_Ch	-1309	+350	1675	Chimpanized Intron
N	ChACPP_Hu	-1309	+350	1659	Humanized Intron
O	HuΔΔACPP_Ch	-1309	+350	1675	*/** Chimpanized Intron
P	ChΔΔACPP_Hu	-1309	+350	1659	*/** Humanized Intron
Q	HuΔΔACPP_X	-1309	+218	1527	*/** No Intron
R	ChΔΔACPP_X	-1309	+218	1527	*/** No Intron
S	HuΔΔACPP_Single	-1309	+333	1642	*/** Single 16bp Region
T	HuΔΔACPP_Double	-1309	+333	1658	*/** Double 16bp Region
U	HuΔΔACPP_Triple	-1309	+333	1674	*/** Triple 16bp Region
V	HuΔΔACPP_Quadruple	-1309	+333	1690	*/** Quadruple 16bp Region
W	pGL4.10	NA	NA	NA	Promoterless Vector

**\*Translation start mutation of ATG to AAG**

**\*\*First Intron 5' Splice Donor Mutation AG/GT to AC/CT**

## 2.3 Results

### 2.3.1 Optimization and Normalization

Most of the biologically relevant results stem from transient transfections of the above-mentioned constructs. However, substantial optimization and validation experiments were first performed. Therefore, I begin this section with a presentation of the results from the optimization of the experimental protocols, validation of controls, and some exploration of the source of experimental variation

#### 2.3.1.1 Transfection Conditions

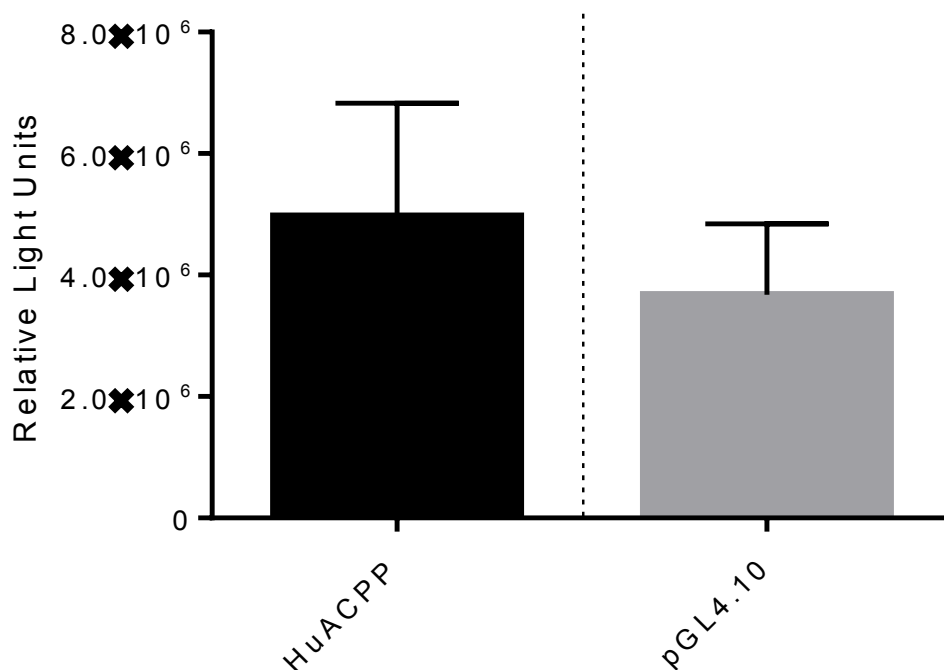
The cell culture assay requires some conditions to be optimized prior to testing the constructs. Two of the conditions were optimized in our lab by Sarah Carnahan-Craig. The first was the optimization of construct transfection parameters using the Fugene® HD (Promega) transfection reagent. The second was to optimize the time from transfection to cell lysis. The optimal ratio of transfection reagent to DNA is 3µl to 1µg of DNA (Craig 2013). The optimal, post transfection, time until cell lysis is 48 hours (Craig 2013).

#### 2.3.1.2 The experimental vector (*HuACPP*) drives expression greater than the promoterless vector (*pGL4.10*).

Using the above transfection conditions, I asked if there was a difference between the signal produced by the experimental vector *HuACPP* and the signal produced by the promoterless vector *pGL4.10*. Each of the experiments included the transfection in

triplicate for each vector. These experiments were repeated on three separate days. The experimental construct (HuACPP) signal is greater, on average, than that of the empty vector (*pGL4.10*) (Figure 2 8).

Note that for these two experiments, 1 $\mu$ g of each vector is added to each well. Because the *pGL4.10* vector is 0.64 times the size of the *HuACPP* vector, the molar amount of *pGL4.10* added to each transfection is 1.6 times greater than the molar amount of *HuACPP*. In all subsequent assays, 1 $\mu$ g of the test vector and 0.64 $\mu$ g of *pGL4.10* are transfected into each well.



**Figure 2-8 *HuACPP* expresses greater than pGL4.10.**

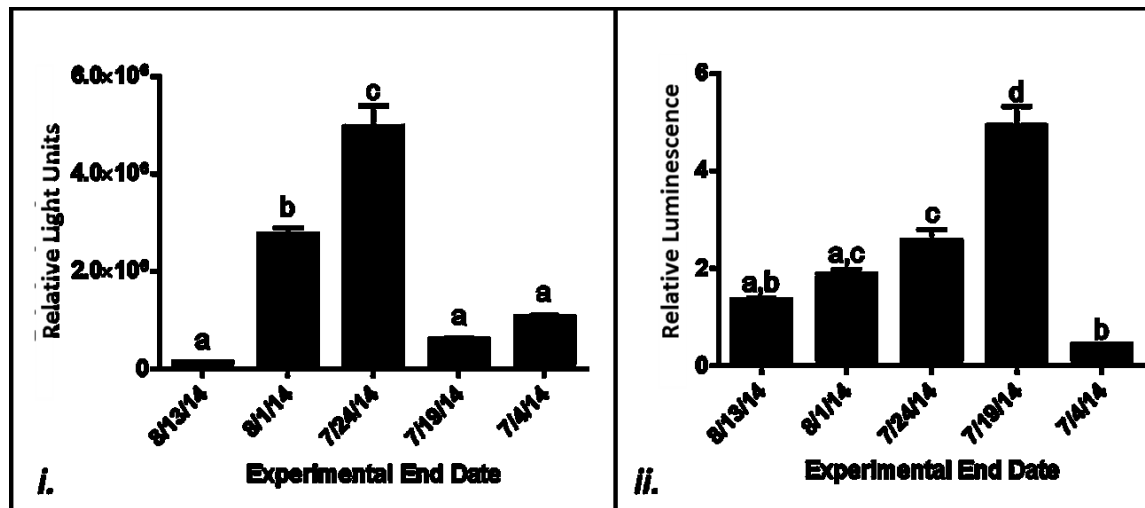


### 2.3.1.3 Normalization of experimental variation

The raw signal of the same construct is not consistent from experiment to experiment, likely due to differences in the state of the cells from day to day or week to week (Figure 2-9i). The most commonly used method for normalization both within and between experiments was to co-transfect a second construct constitutively expressing a different luciferase, such as that of the sea pansy (genus *Renilla*), which because it was internal will also control for transfection efficiency and cell number (Riethoven 2010). Since it was essential to validate any such control prior to its use, we co-transfected LNCaP cells with different concentrations of the pGL4.10 vector (firefly luciferase) with the human *KLK3* core promoter and constant levels of either the pGL4.70 vector (*Renilla* luciferase) (Figure 2-10 i.) or the pGL4.74 vector (*Renilla* luciferase) with the constitutive HSV-TK promoter. Intensity of the *Renilla* signal covaries with the amount of *KLK3* firefly vector transfected, even when small amounts were transfected (Figure 2-10 ii), and with the intensity of the firefly luciferase signal. This covariation has been seen by others in our laboratory when transfecting LNCaP cells, using a wider range of the amount of vector used and with promoters of different human genes (Carnahan-Craig 2013; Das 2014). This “cross-talk” has also been described by others transiently transfecting LNCaP cells (Mulholland *et al.* 2004; Shifera & Hardin 2010).

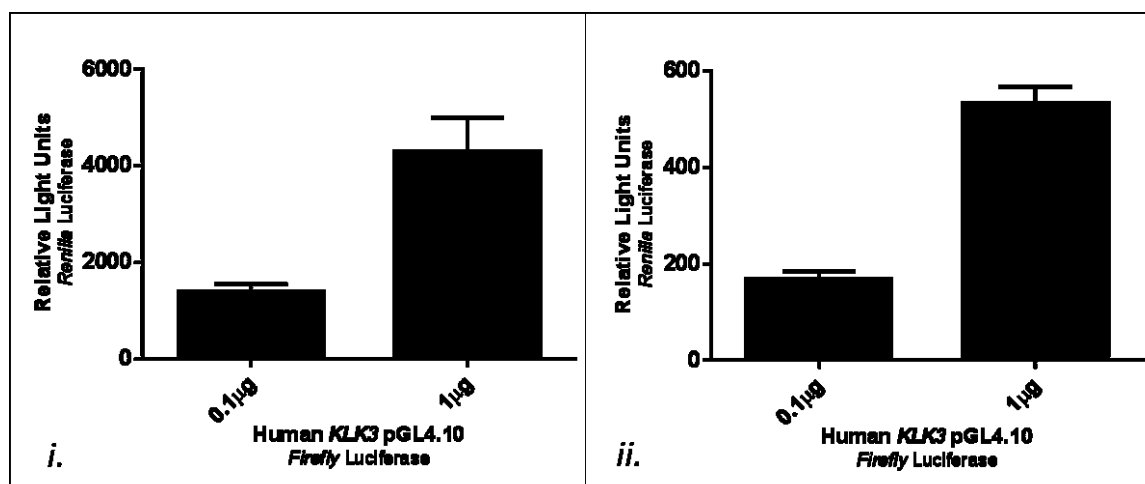
As an alternative to co-transfection, I examined the use of a promoterless pGL4.10 construct for normalization. For this, 1µg of the same experimental construct (*HuΔAACP*) was transfected in triplicate on five separate days; in separate wells of the same plate the molar equivalent (0.64µg) of the promoterless pGL4.10 was similarly transfected in

triplicate. I found this to be somewhat effective (Figure 2-9 *ii*). The average variance by date decreased.



**Figure 2-9 Day to Day Raw and Normalized *HuΔΔACPP* Signal**

The raw signal from *HuΔΔACPP pGL4.10* produced from the same reporter preparations (*i.*) and the normalized signal (*ii.*) from the same experiments. SEM is a measure of variation between 3 replicates for each measure. (One way ANOVA with Tukey's multiple comparison test. Different letters represent significantly different groupings.)



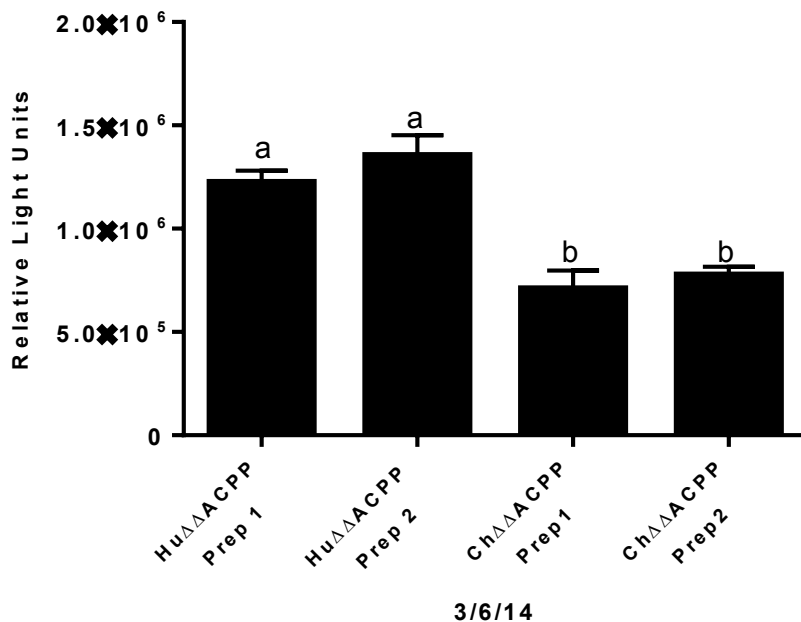
**Figure 2-10 Renilla Signal**

*Renilla* signal when constant amounts of pGL4.70 *Renilla* luciferase vector (0.01 µg (i.) or 0.001 µg (ii.)) are co-transfected with different concentrations human KLK3 pGL4.10 firefly luciferase.

#### 2.3.1.4 Sources of variation

Results from the previous section indicate significant variation when repeating an experiment, especially in the day-to-day variation seen in transfection of the same construct across several weeks. In order to assess other sources of variation we transfected in triplicate two different preparations each of the *HuΔΔACPP* and *ChΔΔACPP* constructs into LNCaP cells on the same day. All wells were seeded from the same source flask. Each ‘preparation’ started from a unique colony on a replicate plate, grown overnight in 50ml LB media on different days and midi-prepped the next day (all midi-preps were done using Qiagen QIAprep Spin Miniprep Kit). The DNA from each prep was quantified, sequence verified, and diluted to the same concentration prior to transfection. As can be seen in Figure 2-11, the variation between the signals from different preps of the same construct is not significantly different, while the difference between constructs is clearly and

significantly different regardless of prep used. Furthermore, the variation among the triplicate transfections is minimal (error bars in Figure 2-11 are SEM). Finally, I tested for variation among duplicate luminometer measurements from the same cell lysate, and found them to be almost identical (data not shown).

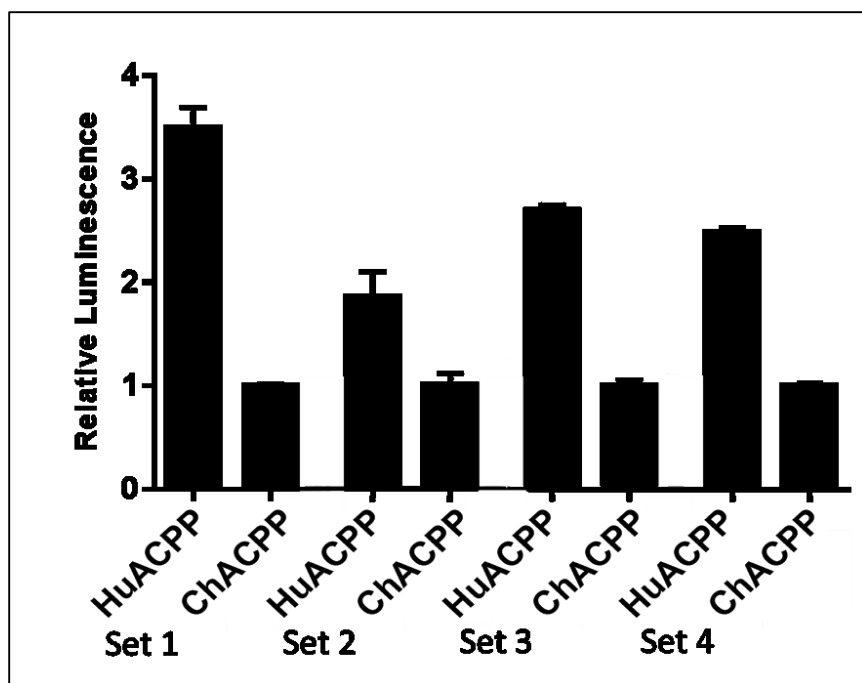


**Figure 2-11 Similar Expression of Different Preparations .**

Different preparations of the reporter constructs expressed on the same day are not significantly different. SEM is a measure of variation between 3 replicates for each measure. (One way ANOVA with Tukey's multiple comparison test.)

Considering these results, I chose not to use a co-transfected *Renilla* luciferase construct to normalize the firefly luciferase values. As an alternative, most of the results shown in the remainder of this chapter are achieved by repeated experiments, where each experiment includes three transfection replicates of each construct (technical replicates rather than

true independent replicates). The average of these technical replicates is normalized by dividing by the average signal from the promoterless pGL4.10 vector, to give the normalized expression for that experiment. Among the experimental replicates, I include at least two independent preparations of the construct except where specified. Finally, I note that an alternative way to present the data is to simply normalize by one of the experimental constructs, as a substitute to an external control. For example, the *HuACPP* and *ChACPP* constructs were transfected in triplicate in four separate experiments run on four separate days, with the value from each transfection replicate divided by that from the average *ChACPP* triplicates (Figure 2-12). The *HuACPP* signal fluctuates from about two to three times greater than that of the *ChACPP*. Though using an experimental result is not an ideal way to normalize data in some respects, it has the advantage of clarity when qualitatively comparing results across multiple experiments, and in requiring no assumptions about the behavior of an external control.



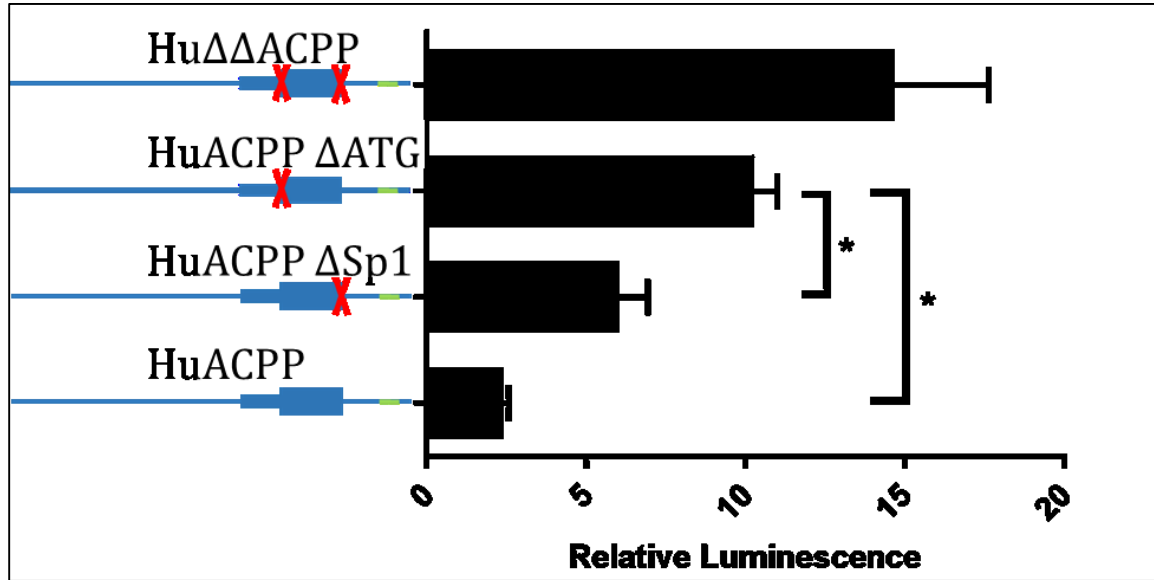
**Figure 2-12 Similar Day to Day Trend in Expression**

Reporter constructs expressed on different days (sets), with each set normalized by the average *ChACPP* signal from that day. SEM is a measure of variation between 3 replicates for each measure.

### 2.3.1.1 Expression increases when the first ACPD exon is deactivated.

Although the difference of expression between *HuACPP* and *ChACPP* is significant (Figure 2-12), the overall signal was low when compared to pGL4.10 (Figure 2-8). This could be due to issues associated with the active 5' exon within the promoter region of the constructs. These issues could include some early out of frame translation, starting at the ACPD translation start site instead of at the luciferase start site, and splicing of the coding region from the mRNA due to active splice donor in the 5' exon. These regions were

independently knocked out of *HuACPP* and transfected in triplicate (Figure 2-13). The elimination of the splicing and translational signals increases overall luciferase expression.



**Figure 2-13 Expression with Step-Wise Deactivation of First Exon**

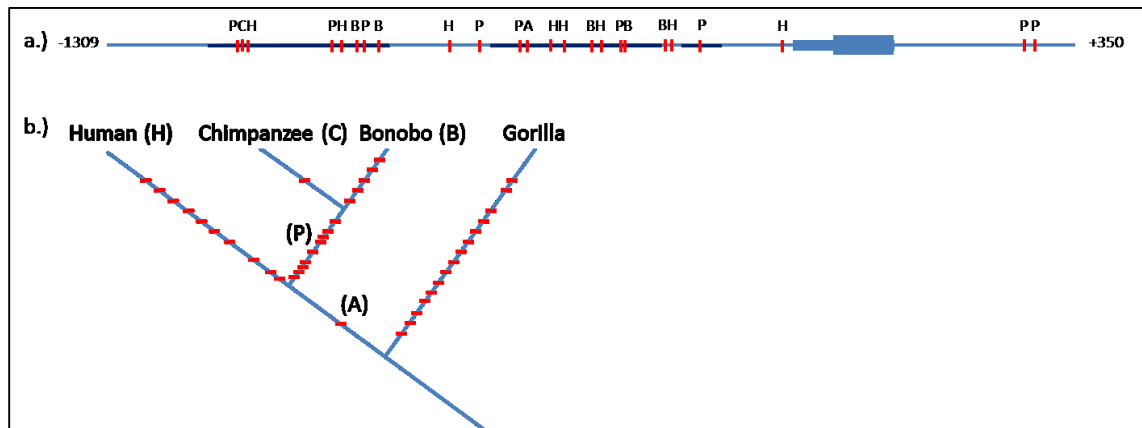
The different human ACPP reporter constructs expressed on the same day. *HuACPPΔATG* signal was significantly greater than both *HuACPP* (\*,  $p=0.0313$ ) and *HuACPPΔSp1* (\*,  $p=0.0208$ ). SEM is a measure of variation between 3 replicates for each measure. (One way ANOVA with Tukey's multiple comparison test.)

### 2.3.2 Sequence differences among the African apes

#### 2.3.2.1 The putative promoter of *ACPP* has species specific differences.

Differences between the *ACPP* putative promoter were determined by aligning the different sequences with Clustal Omega, and using the outgroup orangutan to infer polarity of changes (Figure 2-14). The gorilla branch has 15 single nucleotide differences. The human-*Pan* common ancestor, after the split with gorilla, has one single nucleotide difference, and the *Pan* common ancestor has 9 differences, 8 single nucleotide and one

16bp insertion. Human, Chimpanzee, and Bonobo have 9, 1, and 5 branch specific nucleotide differences respectively.



**Figure 2-14 Mapping of Species Specific Sequence Differences**

Sequence differences along with their physical and phylogenetic location (red) in the *ACPP* putative promoter sequences of human, chimpanzee, bonobo, *Pan* (P), the human-*Pan* ancestor (A) and gorilla. (a.) The relative location of each of the human, chimpanzee, bonobo, and *Pan* differences is marked on the putative promoter, with the repetitive elements denoted by a darker shade of blue. (b.) The phylogenetic origin of each difference is labeled on each branch.

### 2.3.2.2 *Pan* (Chimpanzee and Bonobo) has a 16bp duplication in first intron of *ACPP*.

A 16bp region in the first intron of *ACPP* is tandemly duplicated in chimpanzee and bonobo (Figure 2-15). There are also two additional nucleotide sites within this same region that vary among species, a G/C eight base pairs upstream from the duplication, and a C/T four base pairs into the duplication (Figure 2-15). Outgroup analysis using macaque indicates that the hominoid ancestral nucleotides were G and C respectively. The G/C transversion occurred in the common ancestor of the chimpanzee and bonobo, whereas the C to T transition occurred in the common ancestor of great apes and humans (Figure 2-15).



		10	20	30	40	50	60	70																																																		
Chimpanzee	A	G	T	C	T	G	A	A	G	G	C	A	A	G	C	G	T	C	A	G	T	T	C	A	T	T	C	T	T	G	A	T	T	G	T	T	T	A	T	C	C	T	T	G	A	T	T	G	T	T	C	C	A	G	A	G	A	C
Bonobo	A	G	T	C	T	G	A	A	G	G	C	A	A	G	C	G	T	C	A	G	T	T	C	A	T	T	C	T	T	G	A	T	T	G	T	T	T	A	T	C	C	T	T	G	A	T	T	G	T	T	C	C	A	G	A	G	A	C
Human	A	G	T	C	T	G	A	A	G	G	C	A	A	G	C	G	T	C	A	G	T	T	C	A	T	T	C	T	T	G	A	T	T	G	T	T	T	A	T	C	C	T	T	G	A	T	T	G	T	T	C	C	A	G	A	G	A	C
Gorilla	A	G	T	C	T	G	A	A	G	G	C	A	A	G	C	G	T	C	A	G	T	T	C	A	T	T	C	T	T	G	A	T	T	G	T	T	T	A	T	C	C	T	T	G	A	T	T	G	T	T	C	C	A	G	A	G	A	C
Orangutan	A	G	T	C	T	G	A	A	G	G	C	A	A	G	C	G	T	C	A	G	T	T	C	A	T	T	C	T	T	G	A	T	T	G	T	T	T	A	T	C	C	T	T	G	A	T	T	G	T	T	C	C	A	G	A	G	A	C

**Figure 2-15 The 16bp Duplication**

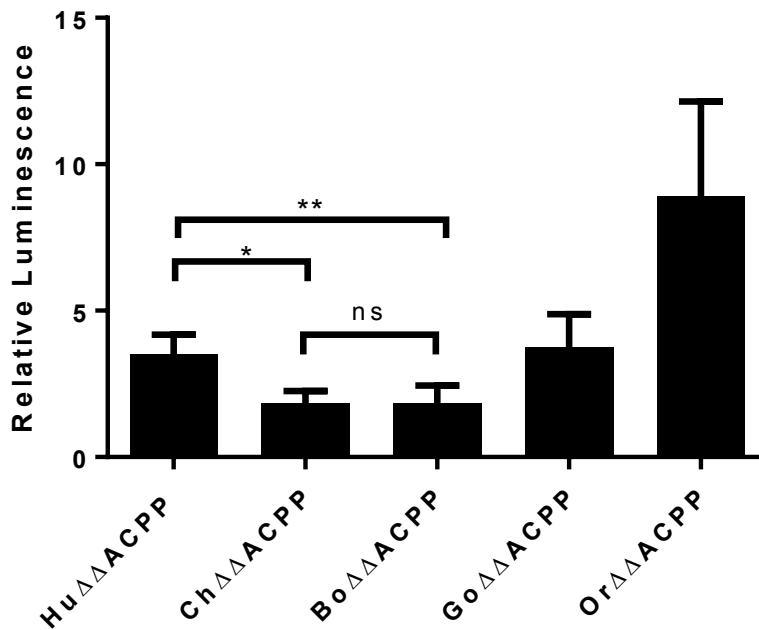
Hominoid sequences aligned to the region surrounding the 16 base pair region of interest in the proximal region of the first intron of *ACPP*. Dashes indicate alignment gaps. The 16bp region (yellow) is tandemly duplicated in *Pan* (green). Other nucleotide differences are indicated in orange. There is one putative SRY binding regions in human, gorilla, and orangutan, and there are two in chimpanzee and bonobo (blue box).

The 16 base pair region, that is duplicated in *Pan*, is located from +290/+306bp from the *ACPP* transcription start site in the human alignment. The 16bp region, as it exist in humans is conserved in the extant African Apes, Neandertal (Vi33.16 and Vi33.25 Sequence Reads, UCSC Genome Browser) and Denisova (High-Coverage Sequence Reads, UCSC Genome Browser). The great ape specific region exists between the +244/+255 androgen dependent, prostate specific GAAAATATGATA-like elements and the +336/+350 ARE. SRY, the male specifying transcription factor, also present in adult male prostate tissue, has a consensus binding site, as predicted with greater than 95% confidence by ConSite (ATTGTTTCC) in the 16bp region , that is duplicated in *Pan* (ATTGTTTTA, and ATTGTTTCC) (Sandelin *et al.* 2004). Whether or not SRY binds, this presents a mechanism by which the 16bp duplication could serve to repress expression of *ACPP*. The duplication could increase the number of binding sites available to repressors. Alternatively, this duplication could insert a novel repressor binding site, or serve to inhibit the binding to an enhancer in the region, by either removing part of the consensus site, or by changing the spacing between two binding sites.

### 2.3.3 Expression differences among constructs and conditions

#### 2.3.3.1 Difference among species ( $\Delta\Delta ACP$ )

To test for species-specific differences of the  $\Delta\Delta ACP$  constructs, I performed seven experiments, each in triplicate, with three separate DNA preparations (Figure 2-16). The *Hu* $\Delta\Delta ACP$  expresses approximately 2 fold greater than either *Ch* $\Delta\Delta ACP$  or *Bo* $\Delta\Delta ACP$  and similar to *Go* $\Delta\Delta ACP$ . Notably, *Ch* $\Delta\Delta ACP$  and *Bo* $\Delta\Delta ACP$  expression profiles are the lowest of any species, and are similar to each other. The increase in *Hu* $\Delta\Delta ACP$  expression over *Ch* $\Delta\Delta ACP$  recapitulates the previously described proteomic data. Within these 7 experiments, as with all following experiments, the trend of lowest expression in *Ch* $\Delta\Delta ACP$  and *Bo* $\Delta\Delta ACP$  compared to all other experimental groups is consistent.



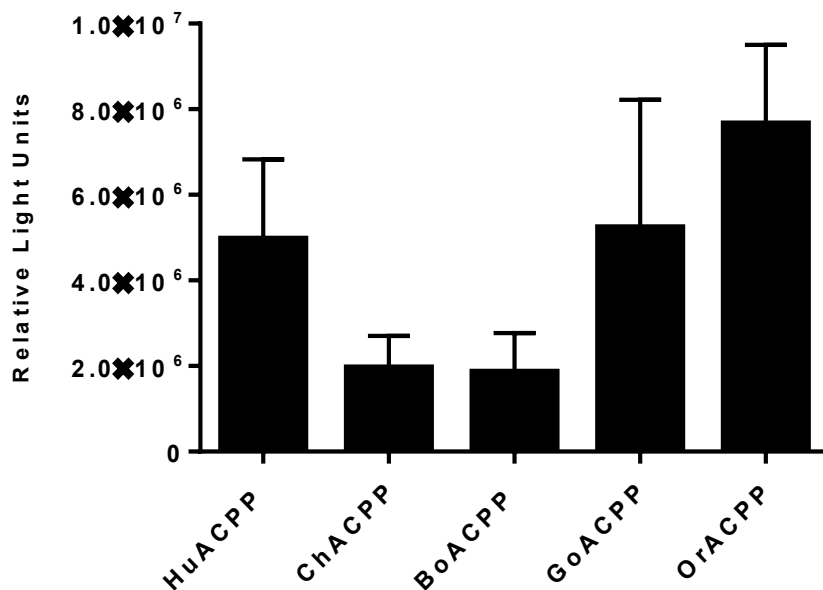
**Figure 2-16 ΔΔACPP Differences between Species (Normalized)**

The different ΔΔACPP reporter constructs from seven different experiments. The *HuΔΔACPP* signal is significantly greater than the *ChΔΔACPP* (\*,  $p=0.0444$ ) and *BoΔΔACPP* signals (\*\*,  $p=0.0062$ ). There is no difference between the *ChΔΔACPP* signal and *BoΔΔACPP* signal (ns,  $p>0.9999$ ). SEM is a measure of variation between 7 experiments for each measure. (One way ANOVA with Tukey's multiple comparison test.)

### 2.3.3.1 Differences among species (ACPP)

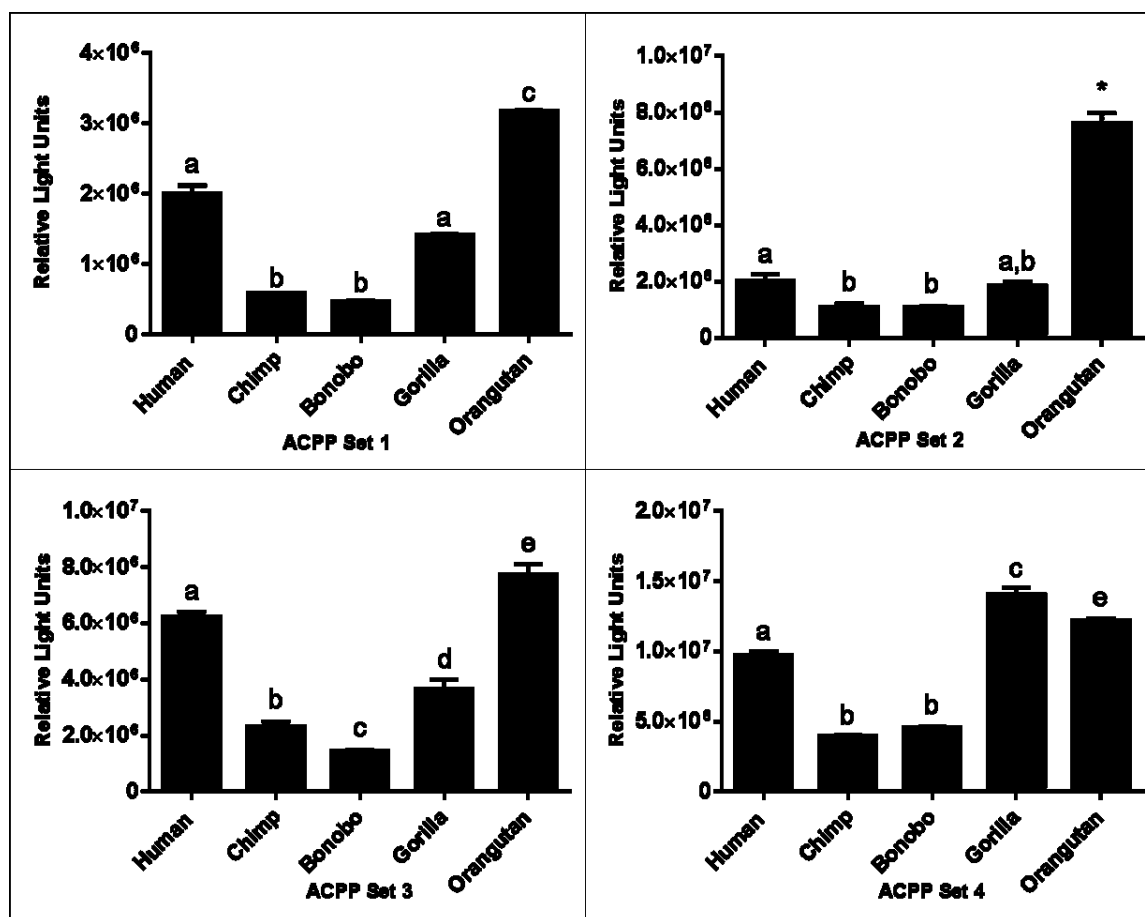
Though the signal is reduced in *ACPP* constructs compared *ΔΔACPP* constructs, I wanted to determine if the removal of the translation start site, and 5' splice donor effected the between species trend, including the increased expression in human over that of chimpanzee. To do this, I performed 4 experiments, each in triplicate, with two separate DNA preparations for each species to test the expression differences between the *ACPP* constructs.

The averaged *ACPP* trends among species are similar to the  $\Delta\Delta ACPP$  trends among species, with the *HuACPP* signal greater than the *ChACPP* and *BoACPP* signal and similar to the *GoACPP* signal, but the trends do not reach significance (Figure 2-17). Even after normalization with the pGL4.10 constructs, the experiment to experiment variability was high so I next present the data as individual experiments (Figure 2-18). In each experiment, *HuACPP* always expresses significantly greater than *ChACPP* and *BoACPP*, when transfection replicates are treated as independent replicates for statistical purposes. *ChACPP* and *BoACPP* always have the lowest signal, and they are not significantly different from each other in three of the four experiments.



**Figure 2-17 *ACPP* Differences between Species (Raw Data)**

The average raw data (non-normalized) from four experiments. The signal variation between experiments increases standard error. No species is significantly different than any other. (One way ANOVA with Tukey's multiple comparison test.)

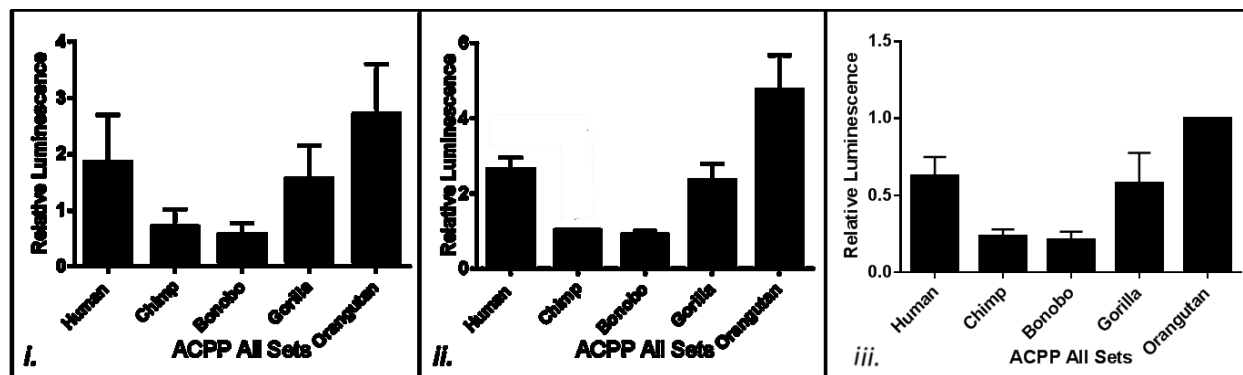


**Figure 2-18 ACPP Differences between Species (Raw Data from Each Day)**

The raw data from each independent experiment. Similar letters indicate similar groups, while different letters indicate significantly different groups ( $p < 0.05$ ). Error bars are SEM derived from 3 transfection replicates of each construct. (One way ANOVA with Tukey's multiple comparison test.)

Because the pGL4.10 signal itself is highly variable (Figure 2-19 *i.*), I alternately explored normalizing the data by the *ChACPP* signal (Figure 2-19 *ii.*) and the *OrACPP* signal (Figure 2-19 *iii.*). Though, normalizing against *ChACPP* is not ideal, as mentioned previously, this has the advantage of clarity when qualitatively comparing results across multiple experiments. Specifically, on average, *HuACPP* expresses with a 2.53 fold increase

over *ChACPP* and *BoACPP*, and that *HuACPP* expressed a signal similar to that of *GoACPP* signal that is on average only 0.05 fold different.



**Figure 2-19 ACPP Differences between Species (Alternative Normalization)**

The average signal normalized against either i.) the promoterless *pGL4.10* luciferase signal ii.) the *ChACPP* luciferase signal or iii.) the *OrACPP* luciferase signal. Error bars are SEM derived from the normalized average of each experiment.

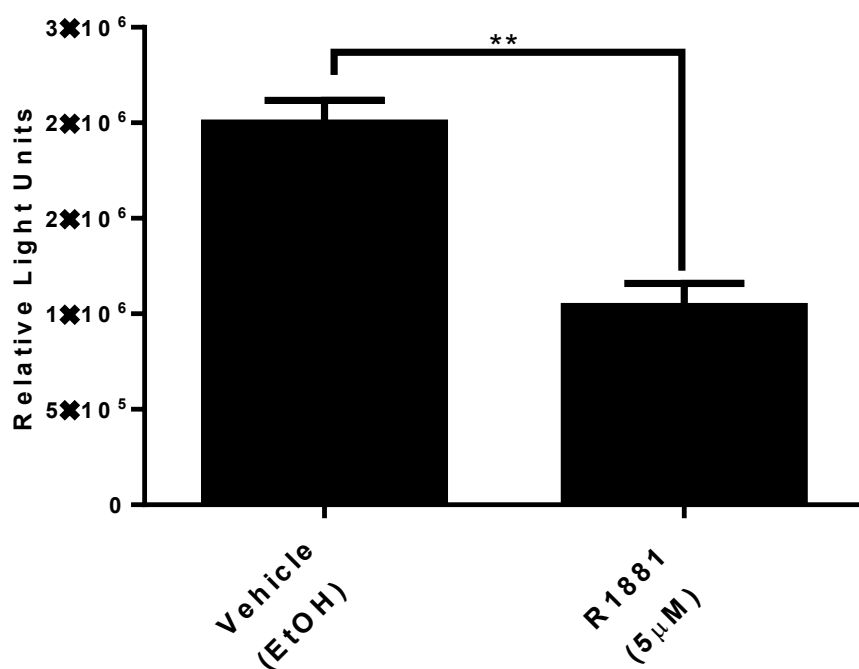
Similar to cotransfection with *pGL4.70* or *pGL4.74*, normalizing the data through the use of in parallel transfection of *ChACPP* or any other experimental construct is not ideal. Instead, I normalize the signal for each experiment against the promoterless *pGL4.10* vector, or I repeat the experiment a minimum of 3 times and show the data without normalization. Though not used in the following experiments, there are some alternative methods which may be useful for normalization of luciferase signal when using the LNCaP cell line. Though not as effective as co-transfection, a control vector such as *pGL4.70* or *pGL4.74* could be transfected in parallel. This does not help when adjusting for transfection efficiency, but like my use of the promoterless *pGL4.10*, in parallel transfection provides a way to normalize the signal between experiments. Unlike my use of the promoterless *pGL4.10*, the use of a *pGL4.74* would provide a signal constitutively driven by the HSV-TK

promoter, but the signal from pGL4.70 and pGL4.74 is derived from *Renilla* luciferase, not firefly luciferase, which possible adds another variable. Ideally, two vectors with different levels of expression should be transfected in parallel, both with consistent signals. Though not used for the purpose of normalization, the pGL4.10 vectors with *HuACPP* and *ChACPP* or *Hu $\Delta\Delta$ ACPP* and *Ch $\Delta\Delta$ ACPP* always inserts were transfected in parallel for all experiments. This recapitulates the trend, already established, for reduced expression of *ChACPP* or *Ch $\Delta\Delta$ ACPP* when compared to either *Hu ACPP* or *Hu $\Delta\Delta$ ACPP*.

### 2.3.3.2 Synthetic androgen (R1881) and *ACPP* Expression

The LNCaP cell line was selected because it is androgen sensitive, and derived from the tissue in which *ACPP* is natively expressed. Because both LNCaP tissue, and *ACPP* expression are androgen sensitive, and because the specific putative promoter used in these experiments had not been tested with synthetic androgen, it is necessary to see how androgen affected expression.

First, I transfected *HuACPP* in triplicate for two treatments: 10nM R1881 and vehicle only. *HuACPP* reporter expression is downregulated by the addition of 10nM R1881 synthetic androgen (Figure 2-20).

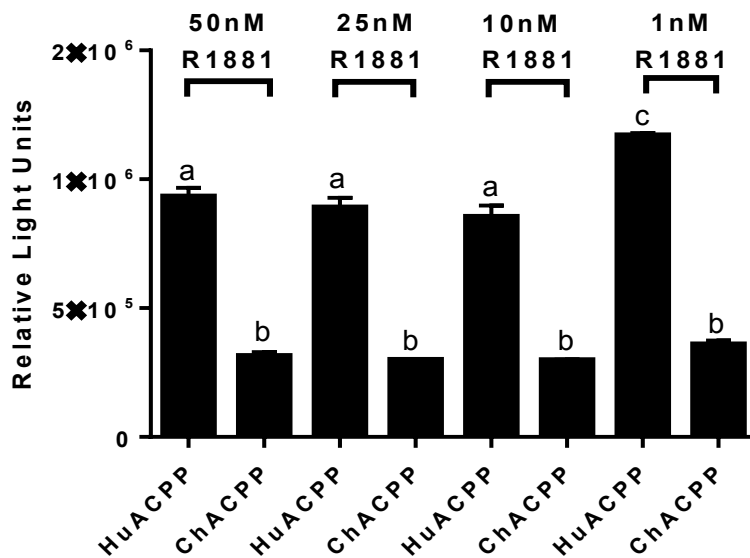


**Figure 2-20 Synthetic Androgen Downregulates Expression**

*HuACPP* reporter expression in the LNCaP cell line. Error bars are SEM derived from 3 transfection replicates of each construct. (p=0.0048; unpaired, 2-tailed t test).

Though 10nM R1881 downregulates expression, I tested to see if this was the lowest concentration of R1881 needed for full repression. To do this, I ran 4 separate treatments in triplicate for both *HuACPP* and *ChACPP*. Both constructs are similarly repressed at 50nM, 25nM and 10nM concentrations of R1881, and both constructs signals increase at 1nM R1881 though only *HuACPP* signal increases significantly (Figure 2-21).

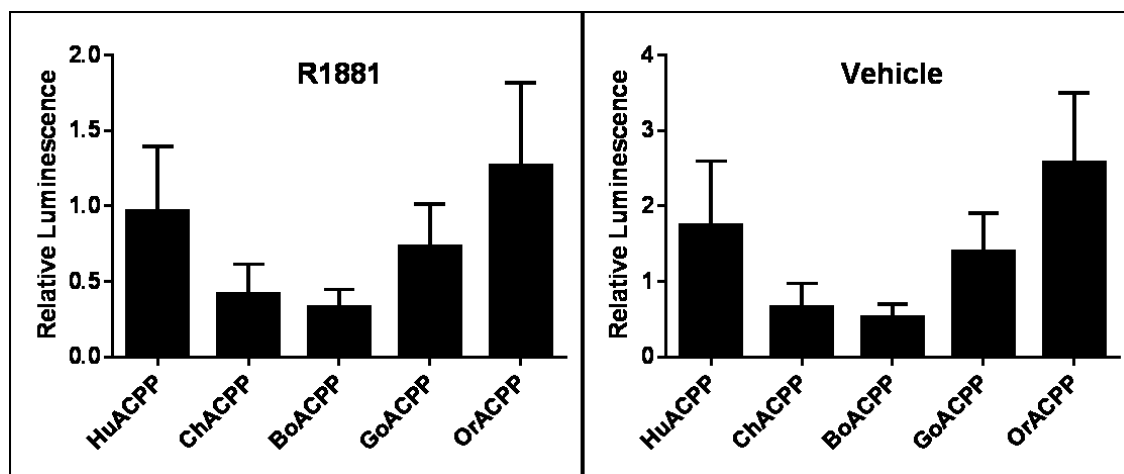




**Figure 2-21 Variable Concentrations of Synthetic Androgen**

*HuACPP* and *ChACPP* raw signal in varied concentrations of R1881 synthetic androgen. Similar letters represent similar groups, different letters represent significantly different groups. The *HuACPP* reporter expression in the 1nM solution is greater significantly greater than in the other concentrations of R1881 ( $p \leq 0.0001$ ). Error bars are SEM derived from 3 transfection replicates of each construct.

The addition of synthetic androgen does not change the trend of *HuACPP* greater than *ChACPP*. The next step was to see if there is a species specific response to synthetic androgen that changes the trend across species. To do this, I transfected every species in triplicate for both the synthetic androgen (R1881) and vehicle (EtOH) treatments. All species reduced expression in androgen (Figure 2-22). To compare the trend between species, I normalized each group by the average *pGL4.10* signal from that group (Figure 2-22). Androgen does not change the trend of human, gorilla, and orangutan *ACPP* reporter constructs expressing at higher levels than either chimpanzee or bonobo.



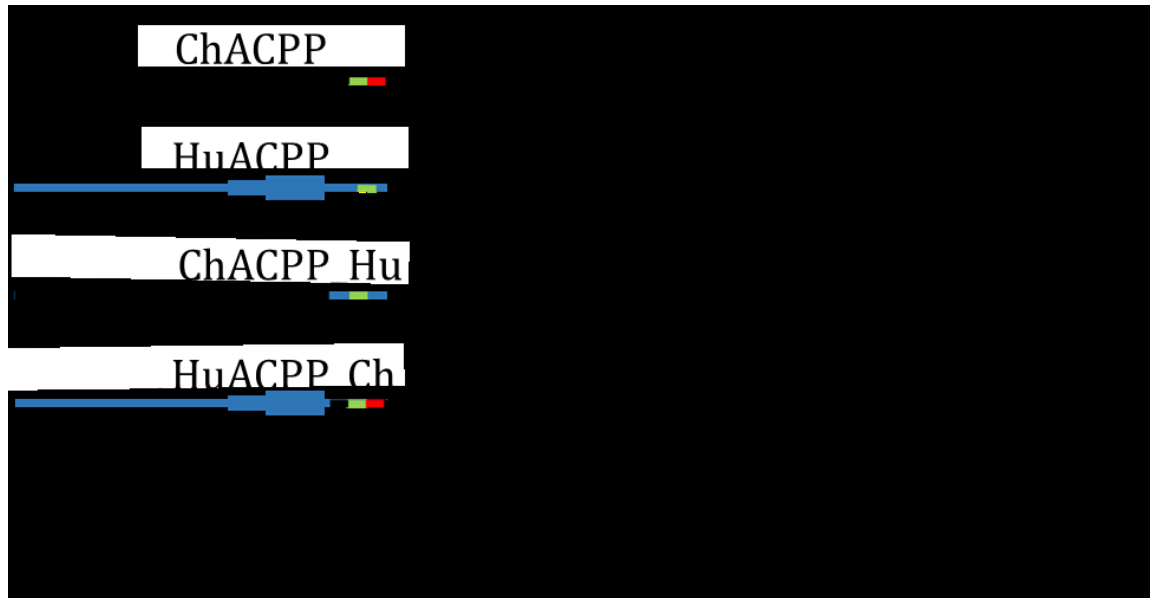
**Figure 2-22 *ACPP* Differences between Species Driven by Synthetic Androgen**

All species *ACPP* signals were normalized by the *pGL4.10* signal from that treatment. SEM is a measure of variation between 3 replicates for each measure after normalization.

### 2.3.3.3 Chimeric *ACPP* (wt and $\Delta\Delta$ )

There are a number of sequence differences between the human and chimpanzee putative promoter region, as shown earlier in Figure 2-14 that could explain the difference between the levels of expression in the two species. When looking at the expression profiles of all of the species together, the most parsimonious change that could explain the reduced expression in both chimpanzee and bonobo would be one that occurred in their common ancestor after the split with human. I choose to look at the largest change, the *Pan* specific 16bp derived duplication that occurs in the first intron of *ACPP* (Figure 2-15). Excluding one single nucleotide difference 8bp upstream of the duplicated region, there are no other differences in the intronic region of the putative promoters of any of the apes. To test if this duplication is responsible for reduced expression in *Pan*, I swapped the 3' intronic regions of the human and chimpanzee *ACPP* reporter construct, giving the *HuACPP*

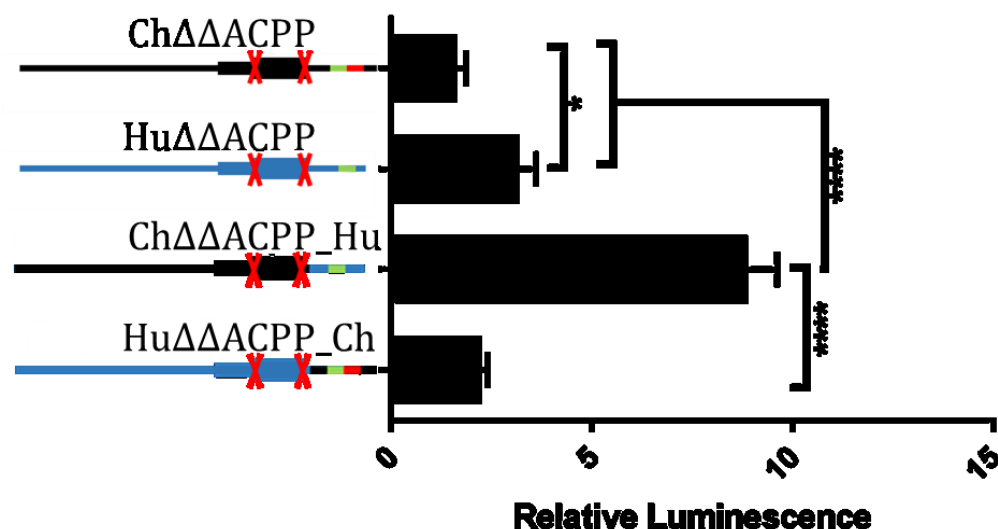
the 16bp duplication (*HuACPP\_CH*), and removing the 16bp duplication from the *ChACPP* (*ChACPP\_Hu*). I then transfected *HuACPP*, *ChACPP*, *HuACPP*, and *ChACPP* in triplicate and normalized using pGL4.10. This experiment was only repeated on one day, with two different preparations of *HuACPP\_CH*, two different preparations of *ChACPP\_Hu*, and a single preparation each of *HuACPP* and *ChACPP*. Notably, the *HuACPP* reporter expressed similar to the *ChACPP\_Hu*, and the *ChACPP* expressed similar to the *HuACPP\_Ch*. This indicates that the 16bp duplication plays a role in the reduced expression of *Pan* over *HuACPP*.



**Figure 2-23 ACP Chimera**

HuACPP\_Ch is significantly less than ChACPP\_Hu (\*\*\*,  $p < 0.0002$ ) and *HuACPP* (\*\*\*,  $p < 0.0002$ ). ChACPP\_Hu is significantly greater than *HuACPP\_Ch* (\*\*\*,  $p < 0.0002$ ) and ChACPP (\*\*\*,  $p < 0.0006$ ). *HuACPP* is significantly greater than *HuACPP\_Ch* (\*\*\*,  $p < 0.0004$ ). The green region of each construct represents the 16bp region of interest, and the red region represents the duplicated 16bp region of interest. The data presented are from one experiment with three technical replicates. SEM is a measure of variation between average signal from each experiment. (One way ANOVA with Tukey's multiple comparison test.)

As shown earlier, the elimination of the splicing and translational start signals increased luciferase expression in all constructs. Therefore, I repeated the above experiments with the *HuΔΔACPP*, *ChΔΔACPP*, and chimeric *ΔΔACPP* constructs. This produced the same trend as before with the *ChΔΔACPP\_Hu* expressing greater than *HuΔΔACPP\_Ch* and *HuΔΔACPP*, and *HuΔΔACPP\_Ch* expressing less than *HuΔΔACPP*. Notably, the *ChΔΔACPP\_Hu* signal is much greater than *HuΔΔACPP*. This again indicates that the 16bp derived duplication significantly represses expression in chimpanzee and bonobo.

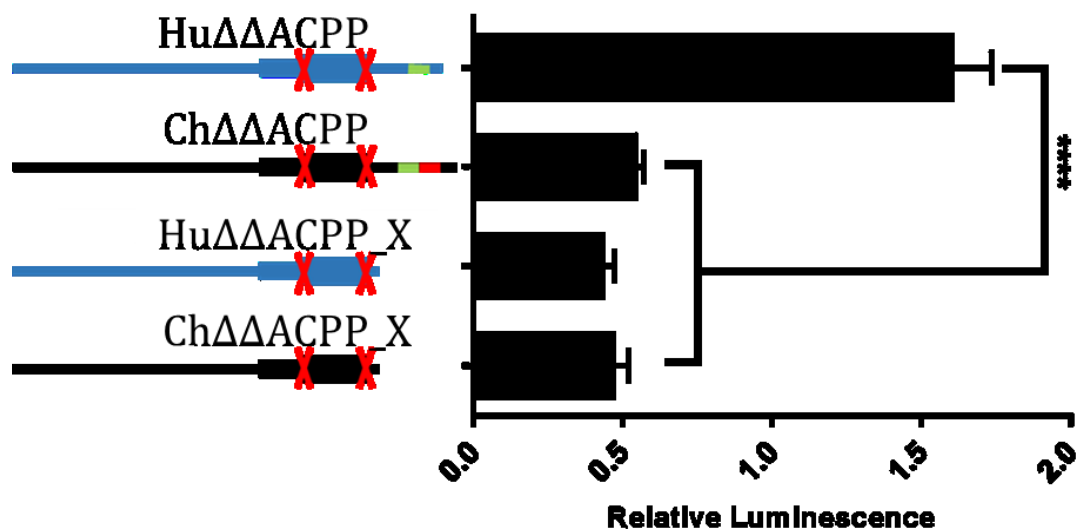


**Figure 2-24 ΔΔACPP Chimera**

These experiments were repeated with two different preparations of each chimeric construct, with each preparation repeated on three separate days, with each transfection performed in triplicate, for a total of 6 different experiments for each chimeric construct, and 3 different experiments for the non chimeric constructs. The signal was normalized against *pGL4.10*. The green region of each construct represents the 16bp region of interest, and the red region represents the duplicated 16bp region of interest. The *ChACPP\_Hu* reporter expression is significantly greater than all other values (\*\*\*\*,  $p < 0.0005$ ) as assessed by one-way ANOVA with Tukey's multiple comparison test. SEM is a measure of variation between average signal from each experiment.

#### 2.3.3.4 Truncated ΔΔACPP

The chimeric *ACPP* results suggest that the species specific expression stems from the intronic differences. To test this, I removed the intron from the *HuΔΔACPP* and *ChΔΔACPP* constructs creating the truncated *HuΔΔACPP\_X* and *ChΔΔACPP\_X* constructs. Remarkably, there is no difference between *HuΔΔACPP\_X*, *ChΔΔACPP\_X*, and *ChΔΔACPP* which indicates that the intronic region of the *ACPP* reporters contains an important region associated with increased expression (Figure 2-15).



**Figure 2-25 ΔΔACPP Truncation**

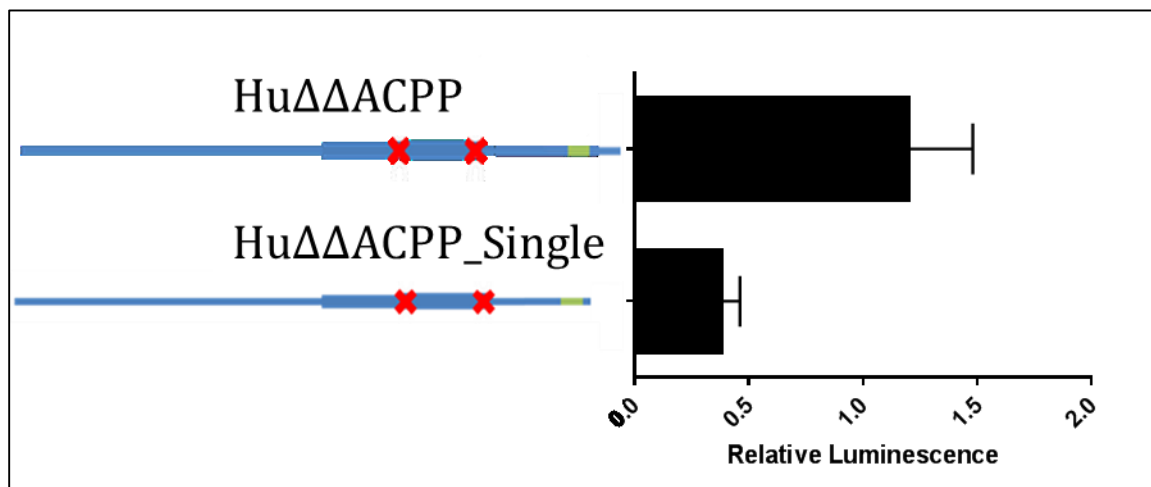
I transfect the truncated constructs with the non-truncated constructs, with two separate preparations made for each truncated construct. The results represent 6 separate experiments, each performed in triplicate, for both *HuΔΔACPP\_X* and *ChΔΔACPP\_X*. The signal was normalized against *pGL4.10*. *HuΔΔACPP* is significantly greater than *ChΔΔACPP* ( $p < 0.0001$ ), the *HuΔΔACPP\_X* ( $p < 0.0001$ ), and the *ChΔΔACPP\_X* ( $p < 0.0001$ ). There is no significant difference between *CHΔΔACPP* and either *HuΔΔACPP\_X* ( $p < 0.5869$ ) or *ChΔΔACPP\_X* ( $p < 0.8315$ ). There is also no significant difference between *HuΔΔACPP\_X* and *ChΔΔACPP\_X* ( $p < 0.9634$ ). The green region of each construct represents the 16bp region of interest, and the red region represents the duplicated 16bp region of interest. SEM is a measure of variation between average signal from each experiment after normalization. (One way ANOVA with Tukey's multiple comparison test)

### 2.3.3.5 A possible trend in Human ΔΔACPP with the addition of one two or three copies of the 16bp region of interest.

The previous sections have shown that the 16bp duplication inhibits expression in a similar manner as deleting the intronic region from the putative *ACPP* promoters. To further investigate the function of this duplication, I designed new human *ACPP* reporter constructs with the 16bp region of interest inserted tandemly either one, two, three, or

four times, as shown in Figure 2-7 and Table 2-1. The constructs were named *Hu $\Delta\Delta$ ACPP\_Single*, *Hu $\Delta\Delta$ ACPP\_Double*, *Hu $\Delta\Delta$ ACPP\_Triple*, and *Hu $\Delta\Delta$ ACPP\_Quadruple*. The primers used to insert these regions during the design of the constructs were long, between 49 and 97 base pairs. To keep these primers from getting too long, I did not include the last 18bp (+332/+350) of the putative promoter into the design.

Testing to see if this small 18bp deletion had an effect on expression, I ran an experiment with *Hu $\Delta\Delta$ ACPP* and *Hu $\Delta\Delta$ ACPP\_Single* with the only difference being the last 18bp. Unexpectedly, *Hu $\Delta\Delta$ ACPP* is an average of 3.2 fold that of the *Hu $\Delta\Delta$ ACPP\_Single* (Figure 2-26). This difference in expression means that it is not useful to directly compare these new constructs to either *Hu $\Delta\Delta$ ACPP* or *Ch $\Delta\Delta$ ACPP*.

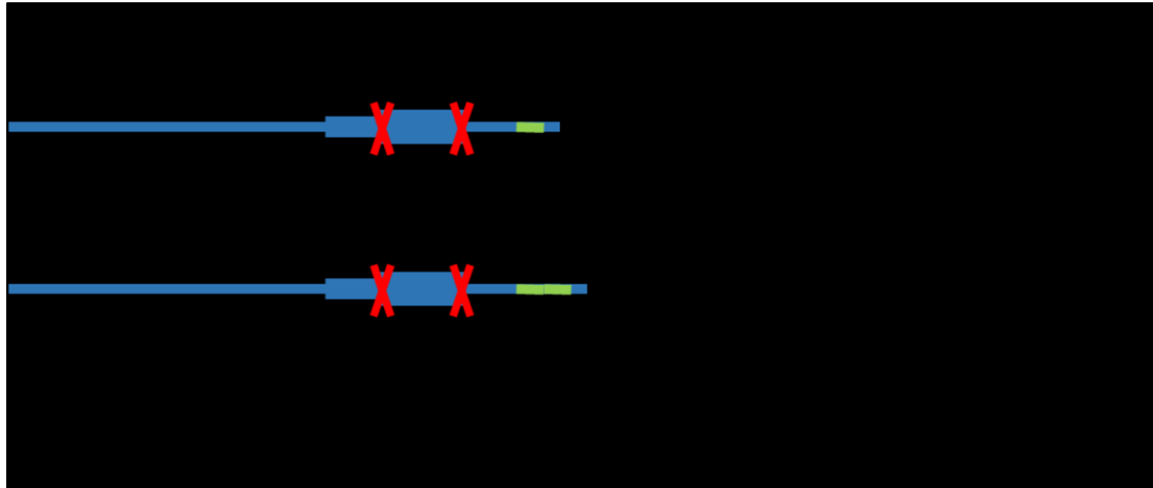


**Figure 2-26 *Hu $\Delta\Delta$ ACPP* versus *Hu $\Delta\Delta$ ACPP\_Single***

The results are a compilation of 6 different experiments, each performed in triplicate, with each construct prepared twice. The signal was normalized using pGL4.10. *Hu $\Delta\Delta$ ACPP\_Single* is significantly less than *Hu $\Delta\Delta$ ACPP* ( $p=0.0430$ ) (Paired, two tailed t test.) The green region of each construct represents the 16bp region of interest. SEM is a measure of variation between average signal from each experiment.

Upon further investigation, the 18bp tail end region +332/+350 contained an androgen response element (ARE) from +336/+350 (Shan *et al.* 2005). This loss of signal in *Hu $\Delta\Delta$ ACPP\_Single* is interesting, but could also be problematic if this region interacts, in a species specific manner, with the 16bp region of interest that is just 32bp upstream of the ARE. To test if the loss of the +332/+350 ARE affected the species specific response, I transfected the unaffected *Hu $\Delta\Delta$ ACPP* and *Ch $\Delta\Delta$ ACPP* constructs (a single preparation of each) along with the affected *Hu $\Delta\Delta$ ACPP\_Single* and *Hu $\Delta\Delta$ ACPP\_Double* constructs (two preparations of each). *Hu $\Delta\Delta$ ACPP\_Single*, again, and *Hu $\Delta\Delta$ ACPP\_Double* signals were much lower than the *Hu $\Delta\Delta$ ACPP*. Remarkably, the *Hu $\Delta\Delta$ ACPP\_Single* expression was greater than the *Hu $\Delta\Delta$ ACPP\_Double* recapitulating the species specific response of *HuACPP* greater than *ChACPP* (Figure 2-27). The loss of the ARE did not change the loss of signal caused by the 16bp duplication, regardless of the overall loss in signal due to the loss of the ARE. On average the *Hu $\Delta\Delta$ ACPP* was 2.6 times greater than the *Ch $\Delta\Delta$ ACPP*, and the *Hu $\Delta\Delta$ ACPP\_Single* was 1.9 times greater than *Hu $\Delta\Delta$ ACPP\_Double*.



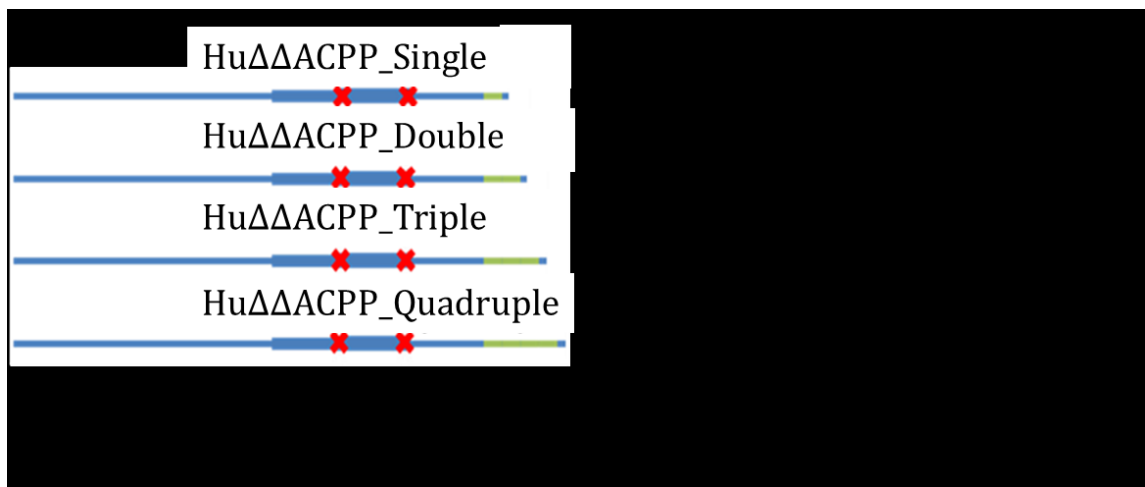


**Figure 2-27 *HuΔΔACPP\_Single* versus *HuΔΔACPP\_Double***

The results are a compilation of 8 different experiments, each performed in triplicate, with each construct prepared twice. The signal was normalized against pGL4.10. SEM is a measure of variation between average signal from each experiment. The green region of each construct represents the 16bp region of interest in the *HuΔΔACPP\_Single* and the duplicated 16bp region of interest in *HuΔΔACPP\_Double*.

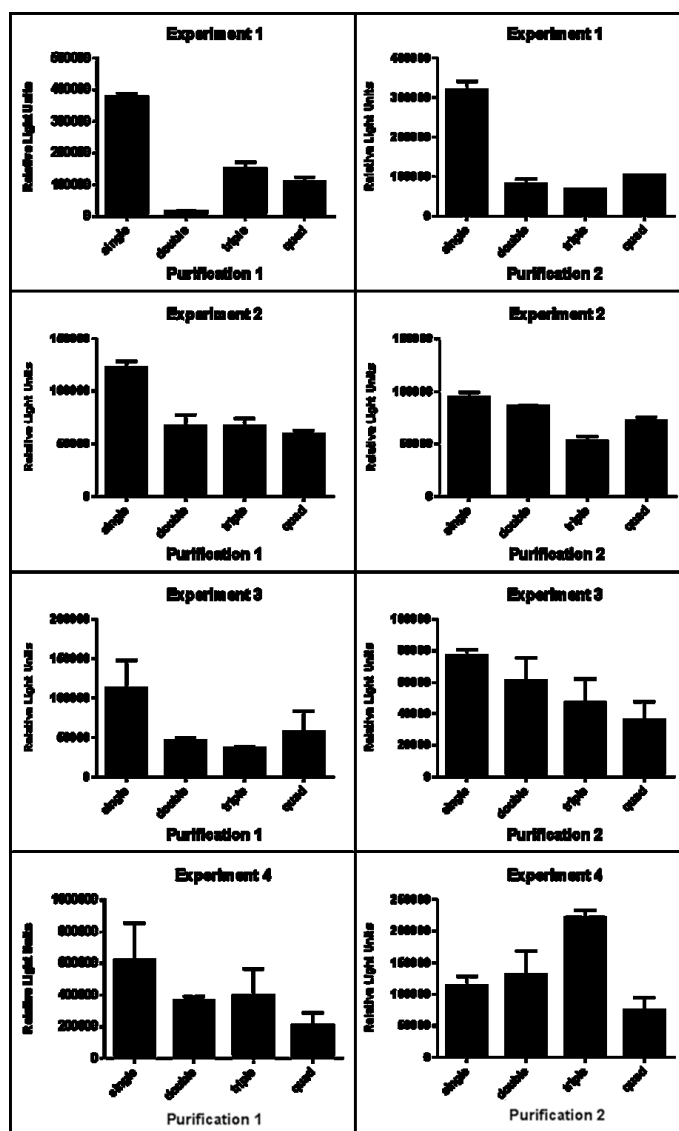
Because *HuΔΔACPP\_Single* and *HuΔΔACPP\_Double* recapitulate the species specific response, and because the 16bp duplication is responsible for the loss of signal in both *ChACPP* and *HuΔΔACPP\_Double*, I next asked how a 16bp triplication or quadruplicating affects expression. To do this, I made two more constructs, *HuΔΔACPP\_Triple* with the 16bp region of interest repeated 3 times, and *HuΔΔACPP\_Quadruple* with the 16bp region of interest repeated 4 times, and transfect alongside *HuΔΔACPP\_Single* and *HuΔΔACPP\_Double*. Because of the large amount of signal fluctuation, I showed the results both combined and normalized by pGL4.10 (Figure 2-28) and the individual raw data (Figure 2-29). When the data is combined, *HuΔΔACPP\_Single* is significantly greater than both *HuΔΔACPP\_Double* and *HuΔΔACPP\_Quadruple*, and *HuΔΔACPP\_Triple* expresses similarly to *HuΔΔACPP\_Double*.

and *HuΔΔACPP\_Quadruple*. The signal varies too much to make any real statement about how the 16bp duplication function. Though the overall decreased signal of *HuΔΔACPP\_Double*, *HuΔΔACPP\_Triple* and *HuΔΔACPP\_Quadruple* leave two options. One is that the region additively decreases transcription, in which case, with each addition of the 16bp region, a larger decrease in signal would be visible. If this is the case, and better results could be obtained, this option would look much like experiment 3, purification 2 in Figure 2-29. The other option is that the 16bp duplication in chimpanzee caused the removal of a transcription factor binding site, in which case there would be no further affect on transcription after duplication, and better results would look something like experiment 3, purification 1 in Figure 2-29.



**Figure 2-28 Combined data for *HuΔΔACPP\_Single/Double/Triple/Quadruple***

The results are a compilation of 8 different experiments, each performed in triplicate, with each construct prepared twice. The signal was normalized against pGL4.10. The green region of each construct represents the 16bp region of interest in the *HuΔΔACPP\_Single*, the duplicated 16bp region of interest in *HuΔΔACPP\_Double*, the 16bp region of interest in triplicate in *HuΔΔACPP\_Triple*, and the 16bp region of interest in quadruplicate in *HuΔΔACPP\_Quadruple*. SEM is a measure of variation between average signal from each experiment.



**Figure 2-29 Independent Experiments of *HuΔΔACPP\_Single/Double/Triple/Quadruple***

Raw data from each of reporter assays. Purification 1 and 2 are experiments with different construct preparations for the same DNA. SEM is a measure of variation between 3 replicates for each measure after normalization.

## 2.4 Discussion

### 2.4.1 *ACPP* patterns of expression in hominids.

#### 2.4.1.1 *ACPP* expression is down-regulated by androgen.

Androgen is known to regulate human *ACPP* expression in the prostate (Lin & Garcia-Arenas 1994; Zelivianski *et al.* 1998). *In vivo* and *In vitro* analysis of *ACPP* expression has revealed that androgen both upregulates and down-regulates expression based on cell density (Lin & Garcia-Arenas 1994), or based on the portion of the regulatory region used as putative promoter regions in reporter assays (Shan *et al.* 1997; Zelivianski *et al.* 2000). Using the largest characterized regulatory region to date, I show that synthetic androgen down-regulates expression of *ACPP* in the LNCaP prostate cell line. This result is similar to known endogenous gene expression previously reported in the same cell line (Waltering *et al.* 2009). I also show that this trend of decreased expression exists across hominids when expressed in LNCaP cells. Because this trend exists similarly across hominids, it may be unnecessary to analyze the species specific responses in light of androgen regulation.

#### 2.4.1.2 *Pan* *ACPP* expression is reduced compared to other hominoids.

The reporter assays show that the *ACPP* regulatory regions drive reduced expression in *Pan* (chimpanzee and bonobo) compared to human, gorilla, and orangutan. The human and gorilla *ACPP* regulatory regions drive expression similarly, while the orangutan region has the highest level of expression.

On average, the human *ACPP* putative promoter drives expression 2 to 3 times greater than *Pan*. This is notable for two reasons. First, the elevated  $d_N/d_S$  value of 1.16 in chimpanzee *ACPP* suggest that positive selection is acting in the coding region specifically along the chimpanzee lineage after its split with humans (Clark & Swanson 2005). Second, and more importantly, these outcomes parallel other results from our laboratory of a several-fold increase in *ACPP* concentration in human seminal plasma compared to chimpanzee seminal plasma (Zielen, Chovanec, & Seaman, *unpublished data*). Pointedly, there is a difference in seminal concentration of *ACPP* that is comparable to the difference seen in promoter-driven expression of *ACPP*, suggesting that sequence differences in the putative promoter may explain most of the *in vivo* differences in protein abundance.

#### 2.4.1.3 A 16bp tandem duplication in the first intron of *ACPP* reduces expression in the chimpanzee-bonobo common ancestor.

Using the orangutan promoter sequence, I found, and phylogenetically placed, 40 differences between the sequences of human, chimpanzee, bonobo, and gorilla. Because I was interested in the *Pan*-specific reduction in expression compared to human and gorilla, I limited my search to the differences that occur on the branch of the chimpanzee-bonobo common ancestor, after the split with human. This branch has 9 differences, 8 single nucleotide, and a single 16bp insert.

Examination of the 16bp insert showed that it was a tandem duplication. Testing this region, using the chimeric chimpanzee (*ChΔΔACPP\_Hu*) and chimeric human (*HuΔΔACPP\_Ch*) constructs revealed that the duplication was sufficient to reduce

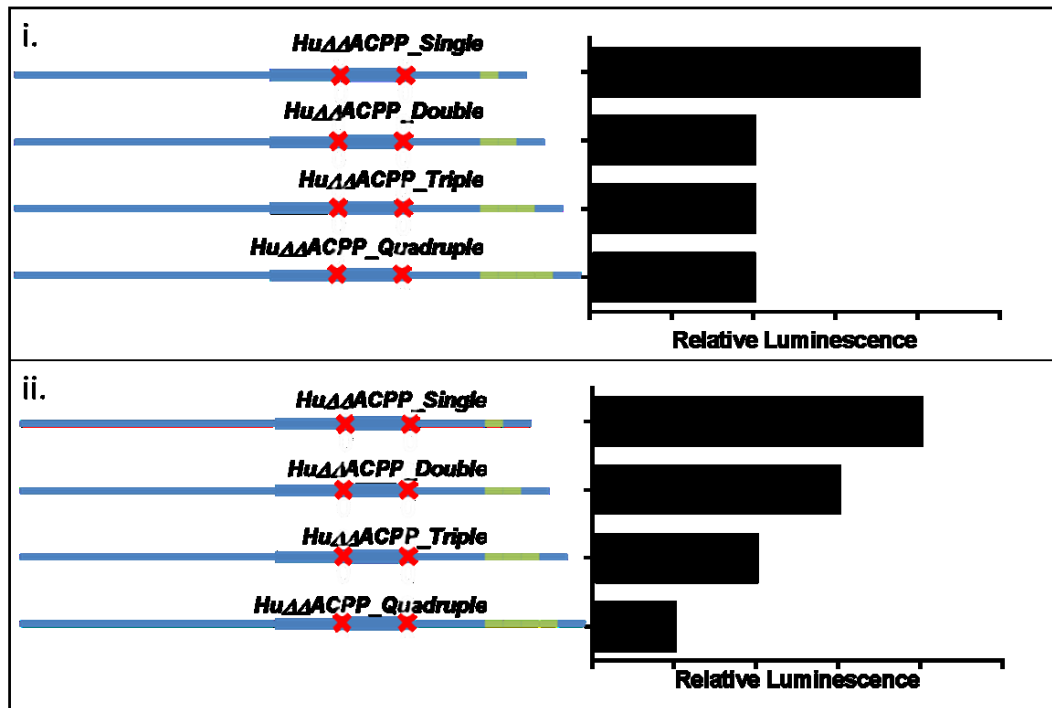
expression in *Pan*. When the duplication was placed into the human construct, the chimeric promoter had reduced expression, and when the single copy of the 16bp region was placed into chimpanzee, expression increased. This confirms that, at least, in part, the 16bp duplication represses expression in chimpanzee.

The region containing one or two copies of the 16bp sequence was removed entirely from the human (*HuΔΔACPP\_X*) and chimpanzee (*ChΔΔACPP\_X*) constructs. These new constructs were used to test if regions outside of the 16bp section were also affecting the level of transcription. Because the only thing that is different in this region is the duplication, any difference that would now exist between the two constructs could be related to any of the other 8 single nucleotide differences on the ancestral chimpanzee-bonobo branch. In this case, both promoters maintained the same level of expression. This suggests that the duplication in chimpanzee is the main difference responsible for a decreased expression.

#### 2.4.1.4 Possible mechanisms causing reduced expression associated with the 16bp duplication.

This duplication could work to reduce expression in different ways. Here, I hypothesize two different mechanisms. First, it disrupts an “up-regulatory” element. Second, the duplication increases the number of sites associated with repression. This could be through the creation of a novel binding site at the intersection of the tandem duplication, or the increase in a pre-existing site or sites already existing in the 16bp region. To test these hypothesis, I created 4 constructs, each with one more 16bp duplication (*HuΔΔACPP\_Single*,

*HuΔΔACPP\_Double*, *HuΔΔACPP\_Triple*, *HuΔΔACPP\_Quadruple*). These constructs were designed to clarify the function of the 16bp region. If the *HuΔΔACPP\_Single* expresses like *HuΔΔACPP*, and the rest of the constructs express like *ChΔΔACPP*, then it suggests that the duplication interrupted a binding site (Figure 2-30,i.). The alternative is that the *HuΔΔACPP\_Single* would again express like *HuΔΔACPP*, but in this case, each subsequent construct would have reduced expression as the number of 16bp regions within that construct increased (Figure 2-30,ii.). For both of these experiments, it is important to consider that the putative promoter of each construct is different in length. It is also possible that the difference in promoter size has an effect on transcription. Hence, qPCR should be used in tandem with the experiment. If the qPCR results are different than the reporter assay, it would suggest that the differences in promoter size have an effect on translation efficiency, and do not represent functional differences driving gene expression.



**Figure 2-30 Possible Outcomes of Increased Copy Numbers of 16bp Region.**

When running the *HuΔΔACPP\_Single*, *HuΔΔACPP\_Double*, *HuΔΔACPP\_Triple*, *HuΔΔACPP\_Quadruple* experiments, the signal was expected to be comparable to the signal produced from the *HuΔΔACPP* and *ChΔΔACPP* constructs. Surprisingly, the signal was reduced, and unfortunately, the experiment with these constructs didn't yield enough signal to eliminate either of the possible mechanisms that could result in reduced expression. Increasing the signal, by reattaching the +332/+350 region may yield enough signal to be able to eliminate one of the mechanisms. Also, adding a strong enhancer to all of the constructs may also help increase signal.



#### 2.4.1.5 The +332/+350 region of the *ACPP* putative promoter drives expression.

Running the *HuΔΔACPP* construct against the *HuΔΔACPP\_Single* construct did reveal that the region contained within the +332/+350 truncation plays a role in driving expression. It is equally interesting that the *HuΔΔACPP\_X*, truncated between +218/+350 *had* reduced expression, compared to *HuΔΔACPP*. Together, these results suggest that both the +218/+331 and +332/+350 play a role in driving expression of the promoter.

Notably, the *HuΔΔACPP\_X* and *ChΔΔACPP\_X* lost the species specific differences associated with the 16bp duplication, and with the truncation of the +218/+331 and +332/+350, they expressed similarly to *ChΔΔACPP*. This is notable because truncation of the +218/+331 and +332/+350 regions yields similar results as adding the 16bp duplication.

#### 2.4.2 A phylogenetic approach to *ACPP* gene expression provides clues to ancestral mating systems.

Many behaviors and their physiological correlates affected by sexual selection are associated with mating system, with broad differences existing among hominoid species. Trying to determine the ancestral behavior of these species is difficult. There is just too much variation in their behaviors and the associated physiological correlates to yield any phylogenetic signal (but see Opie et al. 2012). Using outgroup analysis to determine the behavior of the *Pan*-human common ancestor is not useful because there is no consensus regarding the classification of mating systems, and our nearest living outgroups (*Gorilla*

and *Pongo*) are neither human-like nor chimp-like. The sociosexual behavior of hominids is either too complex, or it evolves too rapidly among hominoids to detect any signal.

Some authors have used physiological correlates of behavior to produce models aimed at describing the evolution of the social behavior of humans. Some assume the *Pan*-human ancestor had a *Pan*-like polygynandrous system with high levels of sperm competition. Starting with traits associated with high levels of sperm competition the transition towards modern humans required such changes as decreased testes size, and a loss of both the copulatory plug and sexual swellings to signal estrus (Lovejoy 2009; Gavrillets 2012). Other authors propose models with a gorilla-like monandry as the initial social system of the *Pan*-human ancestor. These models propose the move to modern *Pan* required an increase in testes size, the formation of a copulatory plug, and the occurrence of sexual swellings. Therefore, the move to modern human, was likely associated with increased paternal care, and reduced male-male aggression (Nakahashi & Horiuchi 2012; Chapais 2013).

A phenotype strongly associated with behavior may be a trait that can be indirectly substituted for behavior. Seminal phenotype is one such phenotype. Being able to recapitulate the ancestral state of the seminal fluid then allows an individual to make well-informed assumptions concerning ancestral behavior based on that seminal phenotype. One can then work forward, to determine on which branches the trait, and thus the behaviors, are derived or ancestral.

As I have shown, the 16bp duplication in chimpanzee results in a derived reduction in *ACPP* expression, matching the *in vivo* quantitative proteomic results. As *ACPP* is implicated in seminal dissolution (Brillard-Bourdet *et al.* 2002), this suggests that plug formation and

loss of seminal liquefaction is a trait derived in *Pan*. This, in turn, suggests that the ancestral mating system was not *Pan*-like in its high levels of polyandry and indicates that models using a gorilla-like monandrous polygyny as the mating system for the last common ancestor are more useful. This in turn suggests models proposing that the transition to modern human required reduced male aggression, both within and between groups, coinciding with increased paternal care, are more useful when considering the evolution of the human mating system. Also, this indicates that the traits associated with increased sperm completion in a multi-male multi-female mating system, like increased testes size and sexual swellings, are derived in chimpanzee.

If the derived reduced expression of ACPP in chimpanzee seminal fluid can be used as an effective proxy for ancestral behavior, it does not invalidate the use of some models asserting a chimp-like polyandrous mating system in the human-chimp last common ancestor. For example, the model proposed by Owen Lovejoy (Lovejoy 2009) is very complex, and considers a large number of variables. For this model, bipedalism arose with and in response to male provisioning of females in trade for copulation. The model suggested by Bernard Chapais (Chapais 2013) suggests that the first step towards social monogamy included weak bonding between different single male polygyny's. Using the logic proposed in Owen Lovejoy's model, male provisioning in exchange for sex could also occur between the weakly bonded single male polygynous groupings proposed by Bernard Chapais. At this point, both models can then still be used to explain a transition to social monogamy.

### 3 Genomic evolution of KLK2 and KLK3 in hominoids

#### 3.1 Introduction

##### 3.1.1 KLK Family

Kallikreins (*KLKs*) are the largest family of contiguously-coded secreted trypsin-like serine proteases in the human genome. They are members of the serine protease family S1 subfamily A of the large chymotrypsin-like serine peptidase clan that have a common structural fold in the absence of high sequence identity (Borgono *et al.* 2004). There are 15 *KLKs* and 1 *KLK* pseudogene coded for in the human 19q13-14 genic locus. They are about 30% to 50% similar in both coding and amino acid sequence with the exception of the notable ~80% similarity between *KLK3* (PSA) and *KLK2*. The *KLK* intronic phases (where the codons start in each exon) are completely conserved, and exon lengths are highly similar to completely conserved between all Kallikreins in human and in mouse (Yousef & Diamandis 2001). The *KLK* origin has been traced back 320mya and has been found in fish, amphibians, and birds, as well as in mammals (Koumandou & Scorilas 2013).

The *KLKs* are split into two groups, the classical and non-classical *KLKs*. These groupings denote the order in which the genes were first described (Lundwall 2013). Though the first *KLK* protein, now known as *KLK1*, was described in the 1930s (Kraut *et al.* 1930) the classical *KLK* genes (*KLK1*, *KLK2*, *KLK3*) weren't described and co-localized until the 1980s (Schedlich *et al.* 1987; Riegman *et al.* 1989a; Riegman *et al.* 1989b). Meanwhile, the non-classical *KLKs* (*KLK4-KLK15*, and the *KLK* pseudogene) were described between

1994 and 2001 (Luo *et al.* 1998; Yousef *et al.* 1998; Yousef *et al.* 1999; Diamandis *et al.* 2000; Yousef *et al.* 2000a; Yousef *et al.* 2000b).

### 3.1.2 KLK Family Protein Activity and Function

Kallikreins are expressed as as prepro-enzymes with inactive or decreased activity, with a signal peptide that is auto-cleaved between the pro-enzymes and secreted. The kallikrein becomes a fully active enzyme when a second peptide is cleaved from the polypeptide (Borgoño *et al.* 2004). They are known to work in cooperation with one another in proteolytic cascade pathways, with one KLK activating another (Pampalakis & Sotiropoulou 2007). For example, KLK5, KLK7 and KLK14 work concordantly in the skin, and there is evidence that KLK2 activates KLK3 through cleavage of the proenzyme in the liquefaction process of seminal plasma (Pampalakis & Sotiropoulou 2007; Lundwall & Brattsand 2008).

The KLKs have a trypsin or chymotrypsin-like substrate affinity (Lawrence *et al.* 2010). This depends on the residue that lies at the base of the substrate-binding pocket. Most KLKs have a trypsin-like affinity with an Asp<sup>183</sup> (chymotrypsin numbering), or in some cases a Glu<sup>183</sup>, in the binding pocket, conferring the ability to cleave at either the Asp or Lys residues (Krem & Di Cera 2001; Lawrence *et al.* 2010). The KLKs with chymotrypsin-like affinity are characterized by Ser<sup>183</sup>, or in some cases an Asn<sup>183</sup> or Gly<sup>183</sup> in the binding pocket conferring the ability to cleave at either Val or Ala residues (Krem & Di Cera 2001; Lawrence *et al.* 2010). The substrate affinity is associated with eight loops on the surface of the protease, another trait that characterizes the KLKs (Debela *et al.* 2008; Lawrence *et al.* 2010).

This group of proteases is characterized by a catalytic triad that is necessary for protease activity (Krem & Di Cera 2001). This triad is conserved along the trypsin-like domain of KLK, and consists of His<sup>57</sup>, Asp<sup>102</sup>, and Ser<sup>195</sup> (Krem & Di Cera 2001). Although some alternative splicing has been observed where one or more of the triad's characteristic elements are missing, it is unknown if this is associated with disease or rather serves some as of yet unknown function, such as one used as an agonist to control the rate of proteolytic activity through competitive binding and rate limiting transcription (Koumandou & Scorilas 2013).

### 3.1.3 *KLK* Genomic Evolution

Through phylogenetic analysis, the progenitor *KLK* has been determined to be *KLK1*. This gene, which was named in order of discovery and not of evolutionary appearance, exists in both amphibian and amniotes. It most likely points to an origin 330mya (Pavlopoulou *et al.* 2010). Excluding lizard (lepidosauria), which has its own derived *KLK* cluster, known as the *KLK* orphans, all species have a similar *KLK* evolutionary relationship. However, not all species have all *KLKs*. Through a series of *KLK1* duplications, *KLK5*, *KLK6*, *KLK10*, *KLK14* and *KLK15* were produced in amniotes, and then continued duplication events produced *KLK7*, *KLK8*, *KLK11* and *KLK13* in mammals.

There are several *KLK* duplications that yielded tissue-specific expression of *KLKs*, and which most likely work in concert with one another. It appears that either *KLK9* produced or was produced by *KLK11* in temporally-concordant tandem duplication events. *KLK4* was produced by a *KLK5* duplication, *KLK12* by a *KLK10* duplication. Finally, *KLK2* was

produced by a *KLK1* duplication in early eutherians, and then *KLK3* appears to have arisen from a *KLK2* duplication in catarrhines. All *KLKs* diverged significantly enough from each other to form monophyletic paralogs, except for the classic *KLKs* (*KLK1*, *KLK2*, and *KLK3*), which form their own clade.

The uninterrupted contiguous arrangement of *KLK* genes is conserved in all of the species which have had their *KLK* locus completely sequenced. This includes the lizard, whose *KLKs*, in spite of being uniquely derived, are also located in the same locus and oriented in an uninterrupted contiguous manner. The duplication events led to divergence and neo-functionalization, or divergence and sub-functionalization of one of the new genes.

#### 3.1.4 *KLK* Enzymatic Action and Evolution

Although many of the physiological and biological functions of the *KLKs* are still being investigated, those functions known to be associated with disease in human have had more attention. It is clear that many of the *KLKs* are closely related but have divergent function, and that many work in concert with one another, operating in signal cascades, activating one another, and working to degrade specific substrates (Borgoño *et al.* 2004; Descargues *et al.* 2005) (Sales *et al.* 2010b).

As discussed earlier, *KLK4* is highly related to *KLK5*, having been produced in from a *KLK5* duplication. They are located immediately next to each other on the *KLK* locus in all known animals that possess these genes, and are co-expressed in cervix and breast tissue (Shaw & Diamandis 2007). *KLK4* is one of the nine widely expressed *KLKs* and can be found in many tissue types (Shaw & Diamandis 2007). Conversely, *KLK5* is restricted to

only a few tissue types (Shaw & Diamandis 2007). KLK4 developed divergent functions from KLK5 after the duplication event, while KLK5's function remained constrained (Pavlopoulou *et al.* 2010). Specifically, KLK4 has been investigated for its association with tooth development and enamel strengthening (Kawasaki *et al.* 2014). Its proteolytic activity targets enamelin, which subsequently allows the reuptake of digested enamel polypeptides, assisting in mineralization and preventing the build-up of brittle enamel on developing teeth (Kawasaki *et al.* 2014). Unlike KLK5, KLK4 is only expressed in animals with teeth, and is known to be inactive or lost in animals that have derived loss of adult tooth development, such as the bird and platypus (Meredith *et al.* 2009) (Kawasaki *et al.* 2014). Interestingly, most species that have lost KLK4 expression have also lost expression of its substrate, but in animals with rudimentary juvenile teeth, like platypus, KLK4 is lost but the substrate, enamelin, is still expressed (Kawasaki *et al.* 2014). This suggest that even though KLK4 is a widely expressed KLK, specific constraints, mainly the presence of permanent teeth, select for the presence of KLK4 activity in a species. This also suggests that other proteases are able to compensate for its loss in the other tissues in which it is expressed.

KLK5 is an active mediator in skin desquamation or peeling. Much more is known about this, and the enzymatic action of the kallikreins in this process, than about other KLK5's functions, because of pathologies related to desquamation (Pavlopoulou *et al.* 2010). KLK5 is expressed in the inactive proKLK5 form. It is either self-activated (Brattsand *et al.* 2005) or activated by matriptase (Sales *et al.* 2010a). In a similar fashion, KLK5 activates proKLK7, proKLK14, and proelastase 2 (Brattsand *et al.* 2005). The active KLK14 then acts



in response, to strengthen the activation of KLK5. This feedback is controlled by the LEKTI inhibitor (Descargues *et al.* 2005). Under- and over-expression of the kallikreins or their inhibitor is associated with multiple tissue pathologies. Some of the substrates that KLK5, KLK7, and KLK14 work on, outside of activation of one another, are respectively desmocollin, desmoglein, and corneodesmosin (Borgono *et al.* 2007). Again, the presence of the KLKs is associated with the presence or function of their substrates, as well as with the order of appearance of the KLK (Pavlopoulou *et al.* 2010). For example, lizards have reduced skin shedding and less permeable skin. KLK5 and KLK7 are not present, but the skin-forming desmocollin and desmoglein are. In contrast, mouse skin is more permeable, and its KLK inhibitor has a decrease in number of inhibitory domains. This suggests that there is an evolutionary relationship between the KLKs and their substrates (Pavlopoulou *et al.* 2010).

### 3.1.5 KLK2 and KLK3

The KLK3 and KLK2 are present in human semen, with KLK3 being a catarrhine-specific protease. KLK3 is also known as prostate-specific antigen, or PSA. These enzymes are associated with viscous seminal dissolution or liquefaction (Borgoño *et al.* 2004). They are expressed in the prostate, where KLK2 self-activates, and then activates the proKLK3 in the seminal fluid (Lovgren *et al.* 1997). Though KLK2 is found outside of catarrhine, it is orthologous to catarrhine KLK3.

The speed of liquefaction has been associated with the presence and strength of sperm competition within a mating system (Dixson & Anderson 2002). Gorillas and gibbons have

low viscosity ejaculates. In gorilla, this is associated with the loss of the coagulatory genes *SEMG1* and *SEMG2*, and in gibbon with the loss of at least *SEMG1* (Jensen-Seaman & Li 2003). TGM4, the SEMG1 crosslinking enzyme, is not active in either species (Clark & Swanson ; Carnahan & Jensen-Seaman). Following the patterns of KLK and substrate evolution described above, this would suggest that the protease or proteases that target SEMG1 and SEMG2 during liquefaction would be lost if they had no specific target

KLK2 and KLK3 have been intensely studied in association with prostate cancer, with increasing levels of blood serum KLK3 (PSA) perceived as a positive clinical indication of prostate cancer (Schieferstein 1999; Loeb & Catalona 2007). Though KLK3 has been studied in association with fertility, there has been no direct link to either blood serum or seminal plasma level and fertility in humans, though there has been a slight positive correlation between reduced sperm motility and increased KLK3, and a negative correlation of KLK3 to fructose concentration (Schieferstein 1999).

Previous research into the evolution of KLK2 suggested that the gene function was lost due to the lack of exons 2, 3, and 4 (Clark & Swanson 2005), or that the whole gene was deleted (Marques *et al.* 2012). While I was working on this project, Marques *et al.* (2012) (Marques *et al.* 2012) published their own findings describing the gain and loss of KLK3 and KLK2 in primates, and describing the gorilla and gibbon *KLK3* and *KLK2* genomic region in greater detail. My work supports their findings in gorilla and gibbon, while providing greater insight into the deletion event and the genomic reorganization in gorilla.

## 3.2 Material and Methods

### 3.2.1 Long Range Fragment Amplification

Long Range PCR amplification was performed in 50µl reactions containing 20ng of genomic DNA from gorilla, 250uM dNTPs, 400µM forward and 400µM of the reverse primers KLK2gene\_LR-F1 and KRSP\_LR\_R1 (Table 5-3), 5µl of TaKaRa LA Taq buffer, 0.5µl TaKaRa LA Taq polymerase, and PCR grade water to 50µl volume. The reaction included an initial 94°C step for 1 minute followed by 35 cycles of 98°C for 10 seconds and 58°C for 10 minutes. After cycling, the reaction was finished at 72°C for 10 minutes and stored at 4°C. Amplified product was run on a 1% agarose gel with an expected fragment size of 10,750bp henceforth known as the “10kb fragment”. The amplification was also performed using genomic DNA from *Homo sapiens*, three *Gorilla gorilla* individuals (Chipua, Fredricka, and Josephine), one *G. beringei* (M’kubwa), *Hylobates leucogenys* (PR598), and *Hylobates syndactulus* (PR1038)

### 3.2.2 Sequencing of 10kb Fragment

All sequencing reactions were run at 20µl using 1µl of BigDye® Terminator v3.1, 4µl of sequencing buffer, 150ng to 250ng of purified product, 3.2pmol of primer, and brought to volume with molecular biology grade water. The reaction included 35 cycles of 96°C for 10 seconds, 50°C for 5 seconds, and 60°C for 4 minutes. After cycling, the reaction was finished at 68°C for 10 minutes, and stored at 4°C.

Sequencing reactions were purified over packed sephadex slurry columns. The columns were packed by adding 550µl of sephadex slurry into well of a 96 well column plate and spun for 3 minutes at 850 x g. The samples were then loaded onto the packed sephadex, and into a 96 well plate by spinning at 850 x g for 4 minutes. The samples were then heated at 98°C for 2 minutes and then cold shocked at 4°C for 2 minutes before analyzing the samples on an Applied BioSystems Avant3130. Sequences were edited and aligned with the help of SeqMan (Lasergene Package, DNASTar).

### 3.2.3 Trace Archives

The NCBI *Gorilla gorilla* WGS trace archives were used to supplement the information obtain from sequencing the KLK long range fragment. The sequences obtained from the 10KB fragment were blasted against the Gorilla gorilla-WGS databases located in the NCBI trace archive. Positive hits (Table 3-1), or the highest scoring sequences and scores with greater than 95% sequence identity were then downloaded, edited, and aligned to the existing consensus using SeqMan (Lasergene Package, DNASTar). The files that were able to be aligned using SeqMan were listed.

**Table 3-1 NCBI *Gorilla gorilla*-WGS database trace files.**

1677261848	2029510746	2051783887	1677261848
2192819663	2033368909	2033145768	2033368909
1677261849	2029729296	2192397050	2035541927
2205028431	2033445149	1679312841	2029813740
1679312841	2029743108	2033145768	2028601745
1688471190	2035541927	1688471190	2267242684
2028601745	2029813740	2033145768	1677261849

### 3.2.4 BAC Library Screen and Sequencing

As described in the following sections, I screened both segments of a male western lowland gorilla (*Gorilla gorilla gorilla*) BAC library (CHORI-255), one segment of a female white-cheeked gibbon (*Nomascus leucogenys*) BAC library (CHORI-271), and one segment of a male human (*Homo sapiens*) BAC library (RPCI-11), all of which were pre-arrayed to nylon membranes. Screening procedures followed protocols described in Ross *et al.* (2001) (Ross *et al.* 2001). BAC libraries and individual clones were obtained from BACPAC Resources (Children's Hospital Oakland; Oakland, CA).

#### 3.2.4.1 Overgo Probe Labeling

The overlapping oligo ("overgo") probes are listed in (Table 5-2) and were designed from non-repetitive conserved regions spanning the 100kb of the hominoid *KLK3* locus. All probes were manufactured with and derived from regions containing 40-60% GC. Each probe set consisted of two 24mer oligonucleotides, each with an 8bp complementary 3'

overlap, designed from one 40bp sequence. 5µl of each oligo (20µM) from a set were mixed, denatured at 80°C for 10 minutes, then annealed at 37°C for 10 minutes before being placed on ice. Oligos were extended and labeled by combining 1µl of BSA (1mg/ml), 2µl of OLB buffer, 4.5µl sterile water, 1µl of the pre-prepared annealed primers, 0.5µl each of alpha-dCTP and alpha-dATP (radioactive), and 0.5µl of Klenow enzyme. After incubation at room temperature for 1 hour, probes were purified over a sephadex column (Nick Column, GE Healthcare).

#### 3.2.4.2 Overgo Hybridization and Screening of Library

The BAC library membranes were pre-warmed in wash buffer (1x SSC, 0.1% SDS) brought to 61°C, then sandwiched between Flow Mesh (Diversified Biotech), and rolled up and placed into a hybridization bottle. Then, 45ml of the preheated, hybridization buffer (Church's hyb buffer: 1% BSA, 1 mM EDTA, 0.5 M NaPO<sub>4</sub> pH 7.2, 7% SDS) was added to bottle and agitated at 61°C for 30 minutes. Probes were denatured at 90°C for 5 minutes, then placed on ice, and added to 5ml of the hybridization buffer. The probe mix was then added to the hybridization tube, and incubated overnight at 61°C. The solution was then emptied from bottle into appropriate radioactive waste container, and the membranes were rinsed 3 times with preheated wash buffer before adding 45ml of wash buffer for agitation for 30 minutes at 61°C. The wash was repeated one more time, then the membranes removed from the tubes, and separated from the mesh. Each membrane was then sealed in plastic wrap, with as little remaining buffer as possible, placed over autoradiographic film in an X-ray cassette for exposure. The film was exposed for 24 to 48 hours at -80°C before developing.

### 3.2.4.3 BAC Culture Purification and Sequencing

Individual BAC clones were grown overnight in 1L of 2X YT culture medium with 12.5µg/µl chloramphenicol. The cultures were then transferred to 500ml Nalgene centrifuge bottles and spun at 6000rcf for 15 minutes at 4°C to pellet cells. BAC DNA was purified with using Clontech's NucleoBond® BAC Maxi kit.

I sequenced the ends of each BAC using SP6 and T7 standard primers. Reactions were run as previously described above, but with a couple of exceptions due to the size of template (BAC DNA > 100kb). Each reaction was run at 20µl using 1µl of BigDye® Terminator v3.1, 4µl of sequencing buffer, 1µg of BAC DNA, and 6.4pmol of primer. The sequencing reaction was run for 75 cycles. Sequencing reactions were purified over packed sephadex slurry columns as described above, before electrophoresis on an Applied BioSystems Avant3130. Sequences were edited and aligned with the help of SeqMan (Lasergene Package, DNASTar). Location of BACs was determined by aligning the end sequences to assembled genomes with BLAT on the UCSC Genome Browser ([www.genome.ucsc.edu](http://www.genome.ucsc.edu)).

### 3.2.5 SeqMan/DNASTar

Sequences were initially aligned using SeqMan, part of the DNASTar computational package. The parameters for assembling the sequences were set using the following options under the Pro Assembler assembly method: Match Size 25, Minimum Match Percentage 80, Match Spacing 150, Minimum Sequence Length 100, Gap Penalty 0.7, and Maximum Mismatch End Bases 15. The assembly fragments were then checked by eye,

using the trace files to insure each base was called correctly, and that gaps were either called or deleted based on trace information. The fragments ended in repetitive sequences that could be sequenced through, by directionally. This bidirectional sequencing allowed for force joining of end. To verify the correct joining of ends, the sequencing primers used to sequence through each end were then used to amplify the region and determine/compare the approximate size of each gap.

### 3.2.6 Protein Structure Prediction

#### 3.2.6.1 Primary Amino Acid Sequence

The gorilla KLK hybrid amino acid sequence was determined by aligning the cDNA from the first four exons of the KLK3 human canonical amino acid sequence and the fifth exon of the KLK2 human canonical amino acid sequence listed with the NCBI accession numbers P07288 and NP\_005542 to the gorilla KLK hybrid consensus sequence using the ClustalW method in MegAlign/DNASTAR. The regions that aligned were then assembled as a coding region in SeqBuilder/DNASTAR, which then translated into an amino acid sequence.

#### 3.2.6.2 Secondary Amino Acid Structure, Active Sites, and Substrate Binding Description.

NCBI's Protein data base for Human KLK3 (P07288.2) contains an extensive list of structural features. Because the majority of the protein was more similar to KLK3 than KLK2, all structural predictions were based on the structural features listed for the



UniProtKB/Swiss-Prot protein, P07288.2 human KLK3. This full 261 amino acid sequence includes the features associated with the un-cleaved signal and pro polypeptide chains.

The gorilla amino acid sequence was determined by aligning the exons of human KLK3 and KLK2 with the nucleotide sequences of the sequences gorilla region. This was used to make cDNA for gorilla. The cDNA was converted to an amino acid sequence, and compared to the human KLK3 and KLK2.

### 3.2.7 Determining KLK Break Point with DnaSP

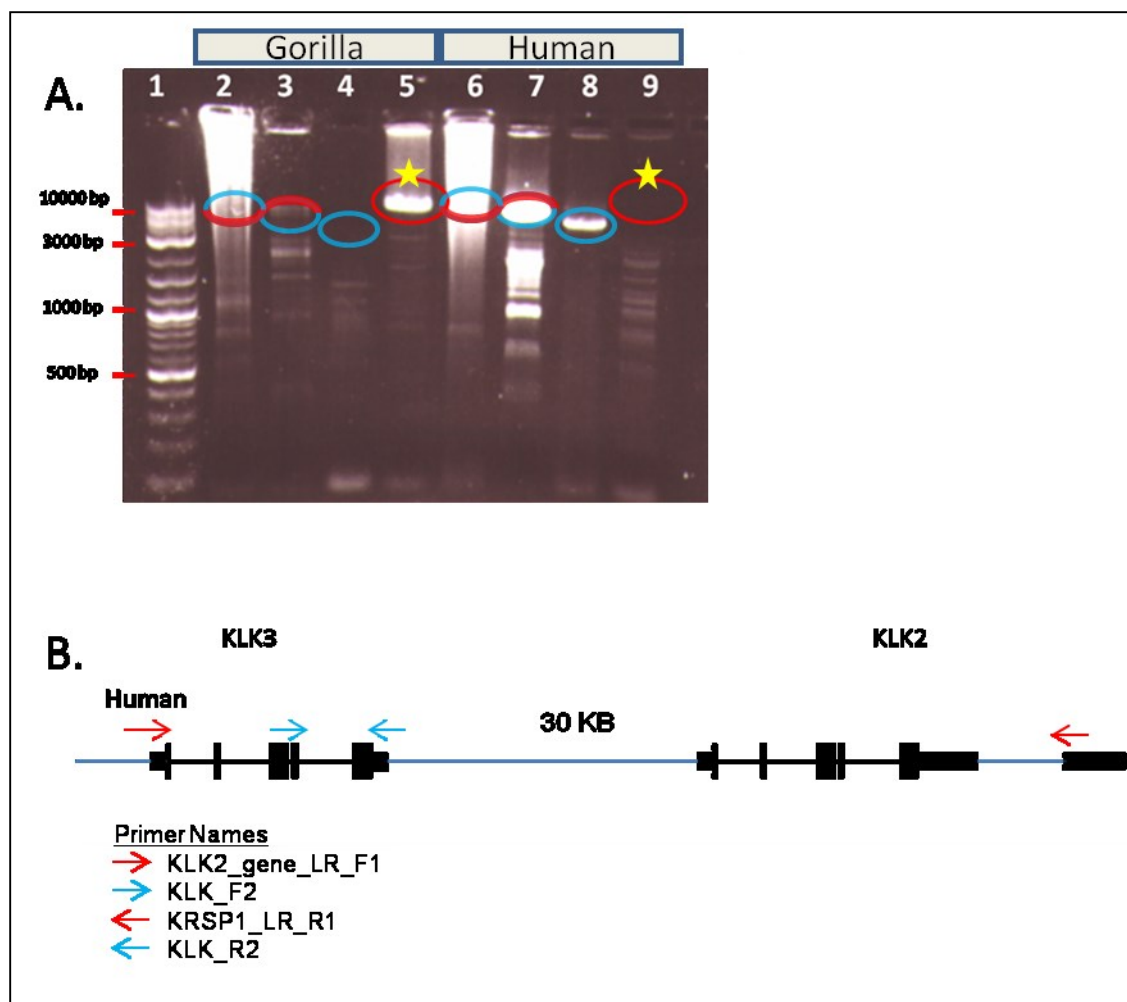
Using Clustal Omega, my Gorilla hybrid genomic sequence was aligned to the UCSC Human *KLK3* and *KLK2* genomic sequences. The Clustal Omega output was saved in an unaligned FASTA file format, with alignment maintained through gaps in each sequence. This file was then opened in DnaSP, a program that analyses aligned DNA sequences. The polymorphism and divergence analysis was run between: human *KLK3* and gorilla *KLK* hybrid, human *KLK2* and gorilla *KLK* hybrid, and Human *KLK3* and Human *KLK2*. The analysis measured divergence using the Jukes and Cantor one-parameter model with a sliding window 1000bp in length, with 10bp steps between each window excluding gaps. The least divergent area between human *KLK3* or *KLK2* and the gorilla *KLK* hybrid were selected as the candidate region.

### 3.3 Results

#### 3.3.1 A 20KB Deletion in the Gorilla *KLK3*/*KLK2* Locus

An approximately 10kb fragment was amplified from gorilla (Chipua) (Figure 3-1 lane 5) genomic DNA but not from human (Figure 3-1 lane 9) genomic DNA using primers developed from the human UCSC assembly. These primers were spaced 30 KB apart in human with the forward primer 700bps upstream from the transcription start site of *KLK3*, and the reverse primer in the *KLK2* proximal end of the *KLKS1* pseudogene (Figure 3-1). In human, the primers would have not been able to amplify the strand due to the 30kb distance, so it is surprising that there was a strong amplified product in gorilla that did not appear in human. Previously, it was thought that only a small deletion had occurred with the gorilla *KLK2* (Clark & Swanson 2005). This band may indicate the possibility that a much larger deletion had occurred. Sequence analysis of the band indicated that it did indeed belong to the *KLK* locus.

When using forward and reverse primers developed within the fourth exon of *KLK3* in combination with the previous primers developed upstream of *KLK3* and within the proximal end of the *KLKS* pseudogene I was able to amplify two fragments in human approximately 10 KB and 3 KB in length (lanes 7 and 8), but I was not able to do the same in gorilla. These results were repeated using genomic template from two other western lowland gorillas (*G. gorilla*) and one eastern lowland gorilla (*G. beringei*).



**Figure 3-1 Gorilla 10KB Fragment**

A) PCR amplifications of *KLK3*/*KLK2* region with stars denoting the 10kb amplification. Ovals indicate areas of interest. As indicated in primer names, the top half of each oval is the color of the forward primer, the bottom half of each oval is the color of the reverse primer. B) The human 30KB region in which the primers were designed.

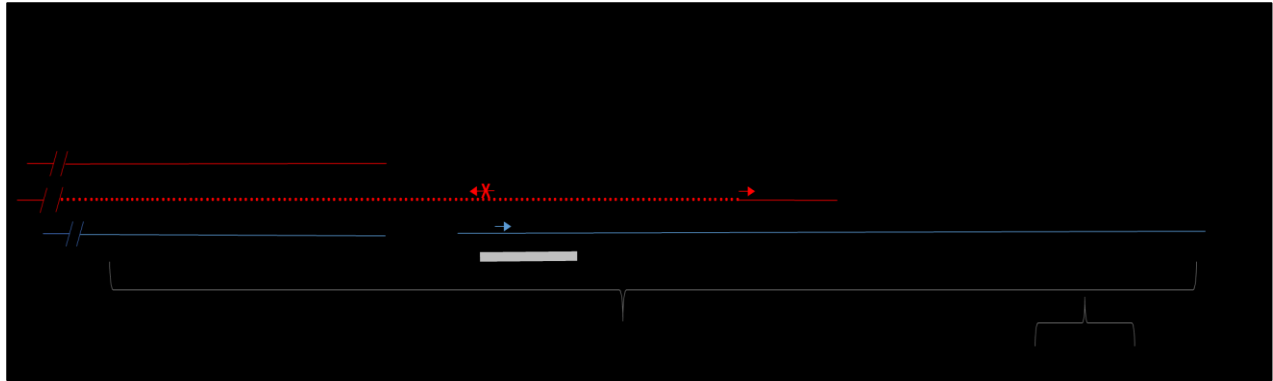
### 3.3.2 Gorilla BAC Library Screening and Sequencing

Due to the repetitive nature of the genomic sequence and the close sequence similarity among homologs at the *KLK* locus, examining this region required larger amplified fragments than those obtainable through traditional PCR. To obtain longer fragments in an

effort to further examine this region, I developed oligonucleotide probes (Table 5-2), within 50kb of the *KLK3/KLK2* locus. These were used to screen BAC libraries from gorilla, gibbon, and human. (Within the rest of this section, to differentiate from the 10KB fragment, I refer to the *KLK3/KLK2* locus in gorilla and gibbon as the chimeric KLK.) Three positive BAC clones were identified in gorilla and three in gibbon (Table 3-2). End-clone sequences mapped these BACs to the *KLK* clusters in gorilla and gibbon (Figure 3-2). The success of the sequencing reactions using the T7 and SP6 BAC-end primers allowed me to map the BAC's to the human public consensus, and determine if regions spanned the chimeric KLK in the comparable gorilla and gibbon genome (Figure 3-2)

**Table 3-2 List of Hybridized BAC Clones**

	<i>G. g. gorilla</i>	<i>N. leucogenys</i> (Gibbon)
<b>Library Name</b>	CHORI-255	CHORI-271
<b>BAC Designation</b>	<u>5G12</u> , <u>42J12</u> , 211L9,	<u>14M16</u> , <u>50N12</u> , 62H2



**Figure 3-2 BAC Mapping to *KLK* Locus of Gorilla and Gibbon**

The gibbon (blue) and gorilla (red) BACs are mapped over the *KLK* locus, with black bands representing genes, and gray bands representing pseudogenes. The BAC ends are demarcated by the applicable T7 or SP6 primer. An unverified BAC region that appears to span the chimeric *KLK* is demarcated by a dotted line. Internal primers that worked are denoted by arrows. The arrow with an X denotes the region and direction that primer (A) was designed from in human.

Regardless of BAC template, only 2 sequencing reactions performed with internal primers were successful (Figure 3-2 BAC Mapping to *KLK* Locus of Gorilla and Gibbon. This includes multiple effort at sequencing the region of the Gorilla (CHORI-255) BAC 5G12, which showed promise as a candidate for sequencing. Though one internal sequencing reaction did work for this BAC (Figure 3-2 BAC Mapping to *KLK* Locus of Gorilla and Gibbon A), it sequenced in an unexpected direction, well away from the region and direction in which it was expected to work (Figure 3-2 BAC Mapping to *KLK* Locus of Gorilla and Gibbon a). Although the BACs spanned the approximately 500KB immediately surrounding the chimeric *KLK* region of interest, no BACs were useful in sequencing across the entire region of interest.

### 3.3.3 Consensus

Regardless of BAC template, only two sequencing reactions performed with internal primers were successful (Figure 3-2 BAC Mapping to *KLK* Locus of Gorilla and Gibbon. This includes multiple efforts at sequencing the region of the gorilla (CHORI-255) BAC 5G12, that showed promise as a candidate. Though one internal sequencing reaction did work for this BAC (Figure 3-2 BAC Mapping to *KLK* Locus of Gorilla and Gibbon A), it sequenced in an unexpected direction, well away from the region and direction in which it was expected to work (Figure 3-2 BAC Mapping to *KLK* Locus of Gorilla and Gibbon a). Although the BACs spanned the approximately 500KB immediately surrounding the chimeric *KLK* region of interest, only one, gorilla CHORI-255 42J12 extended into the first 100bp of the chimeric *KLK* UTR.

Because screening of the BAC library did not yield fragments that extended across the *KLK3/KLK2* region of interest in gorilla, the complete sequence of the *KLK* region was determined by compiling 40 bi-directional sequencing reactions using the 10KB amplified fragment as template, as well as using 24 bi-directional sequences from the trace archives (Table 3-1). The information was aligned, and forms the consensus (Figure 5-3). This information adds to the public assembly by filling in 3 gaps of 271, 1105, and 230 base pairs denoted at chr19:48224088-48224361, chr19:48225553-48225652, and chr19:48227833-48229621 of the 2011 UCSC *Gorilla gorilla gorilla* draft assembly. The first gap occurs within the first intron of a region homologous to human *KLK3*. The second gap occurs in a region that is homologous to the fourth intron of human *KLK3* on the centromeric end, and homologous to the fourth intron of human *KLK2* on the telomeric

end. The last gap in the public assembly occurs between the region homologous to *KLK2* and the *KLKS1* pseudo gene. There are also multiple differences between my consensus sequence and the UCSC gorilla assembly, some of which are possible SNPs. There are two differences between the public assembly and my assembly that occur within my proposed coding region of a novel gene henceforth to be called the gorilla chimeric *KLK*. One of these is a synonymous nucleotide difference in the first exon, and one is a non-synonymous difference in the fifth exon (Figure 3-3).



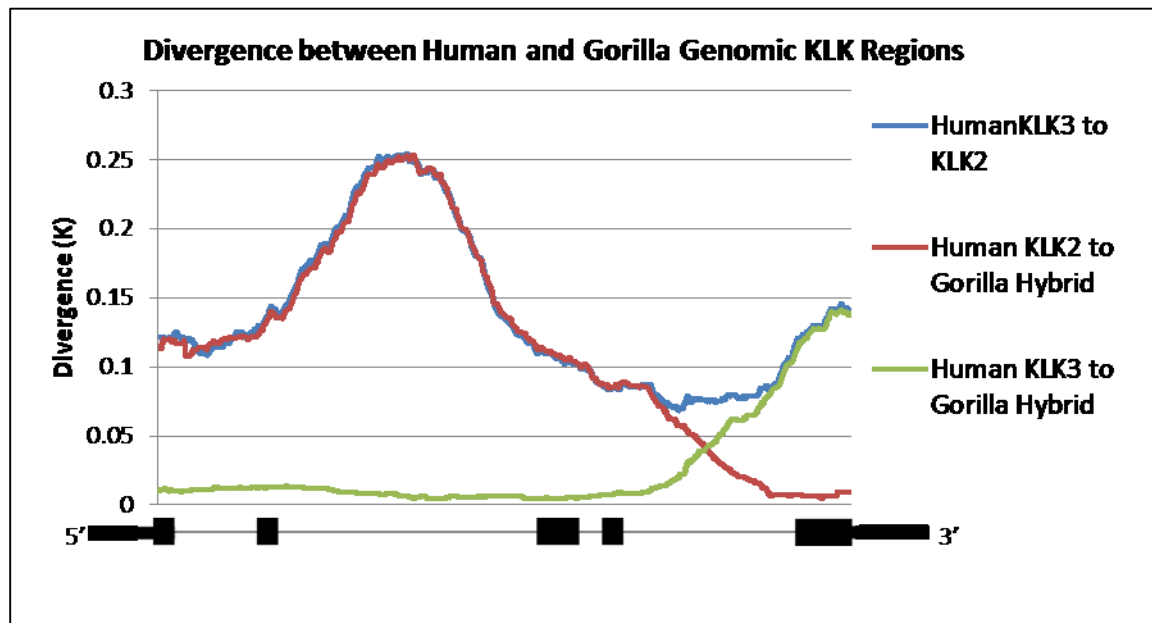
**Figure 3-3 The gorilla chimeric *KLK***

Above is a visual representation of my assembly of the gorilla chimeric *KLK* into the first exon of the pseudo-gene *KLKP1*. Blue lines are differences between public assembly and my assembly. Green lines denote nucleotides that exist in public assembly that do not exist in my assembly. Red denotes synonymous differences, and black denotes non-synonymous differences. Green denotes nucleotides that exist in public assembly that do not exist in mine.

### 3.3.4 Deletion Occurred Between the Fourth Exon of *KLK3* and Fifth Exon of *KLK2*

To determine where the deletion occurred I used DnaSP to measure the divergence between 3 separate sequences. The analysis shows that human *KLK3* and the gorilla chimeric *KLK* are less than 3% divergent through the first four exons of both genes. Human

*KLK2* and the gorilla chimeric *KLK* are similarly divergent starting around the fifth exon of both genes. When comparing both human *KLK3* and human *KLK2*, the region within the fourth intron is the most conserved. The candidate region in which the deletion most likely occurred is within the 1000bp window between the human *KLK2* to gorilla hybrid and human *KLK3* to gorilla hybrid intersection (Figure 3-4).



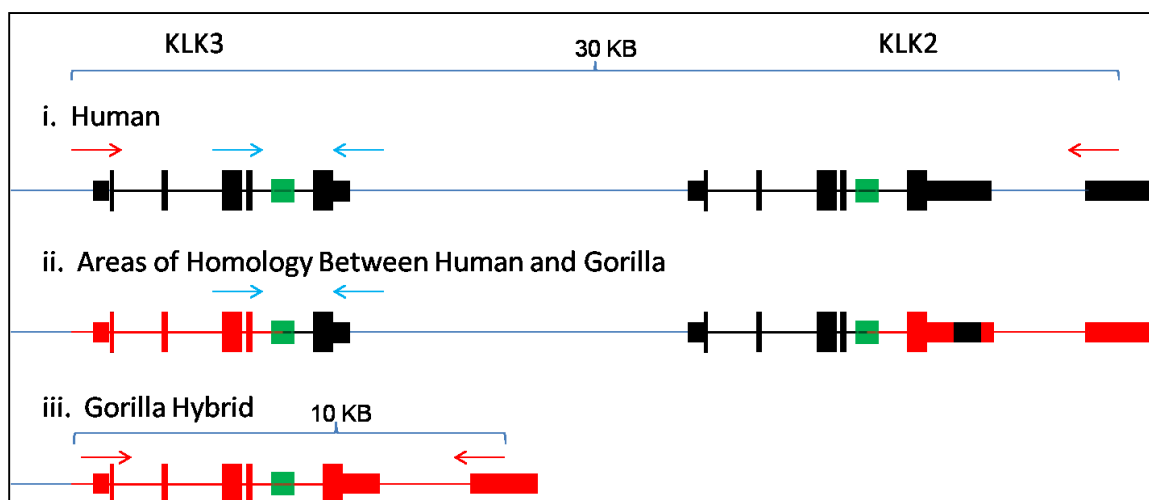
**Figure 3-4 *KLK3*, *KLK2*, and gorilla chimeric *KLK* divergence.**

The Jukes-Cantor divergence is calculated every 10bp within a 1000bp window where K is the average measure of 100 windows of 1000bp that cross each point and started 1000bp from first and last intron. With the x axis spanning about 4700bp from the 5' start codon of the first exon to the 3' stop codon of the last exon of the homologous regions of each *KLK* gene.

The region of the fourth intron of each gene contains a homologous LINE of repeat type L2a, family L2. These LINEs stem from the ancestral duplication of *KLK2* that produced *KLK3* (Pavlopoulou *et al.* 2010). The deletion break point occurred within this LINE (Figure 3-5 ). Aligning the human *KLK* region (Figure 3-5i) with the gorilla consensus gives us



areas of homology between human and gorilla genes (Figure 3-5ii). Because the entire Gorilla consensus matches to human, it is possible to use the human coding regions for each to infer the novel chimeric gorilla *KLK* gene makeup (Figure 3-5).



**Figure 3-5 Homology between human 30KB *KLK* region and the 10KB gorilla chimeric *KLK* region.**

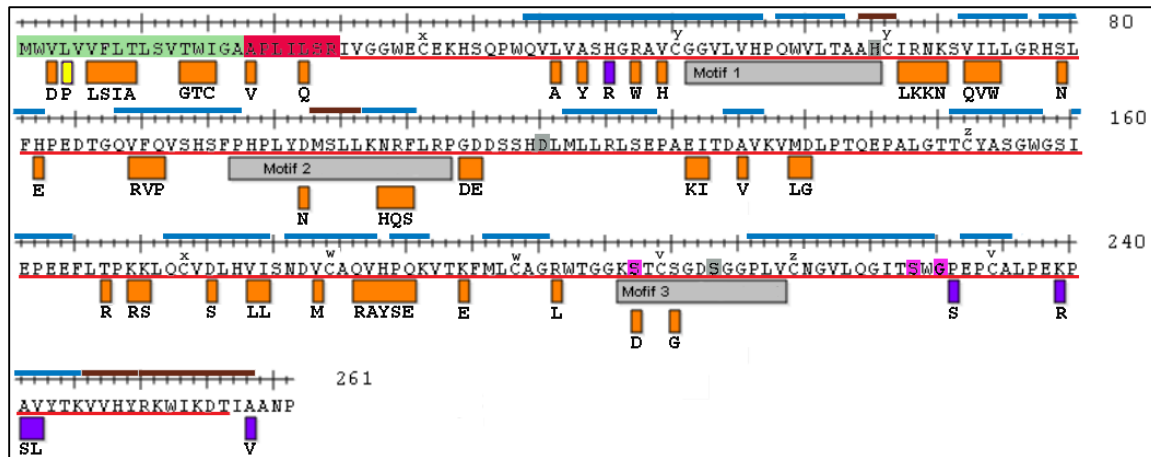
The human *KLK3* and *KLK2* region from UCSC, representing the gorilla ancestral sequence (i) aligned with the gorilla consensus to show the most likely regions of homology (red) between gorilla and human (ii). Two major deletions in the gorilla ancestral *KLK* region (black area of ii) led to the current gorilla region (iii). The green boxes show the arrangement of the LINEs of repeat type L2a, family L2 that are associated with the largest deletion. The arrows represent the primers from Figure 3-1.

### 3.3.5 Resulting Gene

#### 3.3.5.1 Coding Sequence and Catalytic Motifs

To predict the most likely coding sequence of the gorilla *KLK* chimeric gene, I aligned my sequence with the homologous coding regions from human *KLK3* and *KLK2*. Because of

the large amount of homology between the hominoid ancestral *KLK3* and *KLK2* the final proposed 261 amino acid chimeric gorilla preproKLK sequence shares 202 amino acids with both ancestral proteins with 53 amino acids unique to the ancestral *KLK3* and with 6 amino acids unique to the ancestral *KLK2* (Figure 3-6).



**Figure 3-6 The proposed amino acid composition and structure of the gorilla chimeric KLK.**

The gorilla chimeric KLK prepro-protein has 3 catalytic motifs (gray bars) that mark the active chimeric KLK (red line under amino acids) as trypsin-like serine protease. The amino acids specific to ancestral *KLK3* are denoted by orange bars with *KLK2* amino acids notated below. The amino acids specific to *KLK2* are denoted with a purple bar with *KLK3* amino acids notated below. The amino acid specific to the gorilla chimeric KLK and also to the ancestral *KLK3* are marked with a yellow bar with the human *KLK3* amino acid noted below. The signal peptide is highlighted in green and the zymogen contains the peptides highlighted in red. Disulfide bonds are formed between points w,x,y,z. The bars above the gorilla chimeric KLK amino acids sequence denote the regions homologous to human *KLK3* that are beta-strands (light blue) or helical regions (brown). Amino acids highlighted in grey are the catalytic triad or active sites, and in purple are substrate binding sites.

As expected, the N terminal region that corresponds to the first four exons of the proposed gorilla chimeric KLK matches the human *KLK3*, while the C terminal end that

corresponds to the fifth exon of the gorilla chimeric KLK matches the human KLK2.

Surprisingly, two amino acids, Leu<sup>4</sup> and His<sup>45</sup>, do not follow this pattern. The Leu<sup>4</sup> does not match human KLK3 because of a human derived change. At this codon position, human a T to C transition changed the codon from CTG to CCG. This amino acid matches the ancestral KLK3 but not human KLK3 (Figure 3-6). His<sup>45</sup> matches the ancestral KLK2, not the ancestral KLK3. This is due to two nucleotide changes at the second and third coding position resulting in the non-synonymous codon change of CGT to CAC.

The first KLK catalytic motif in the gorilla chimeric KLK, spanning amino acids 51 through 65, is a 100% match to both ancestral KLK3 and KLK2 motifs. The second and third KLK catalytic motifs in gorilla, spanning amino acids 97-113 and 206-218 are a 100% match to the ancestral KLK3 motifs, but differ from the ancestral KLK2 catalytic motif at amino acids 108-110, 207 and 210 (Figure 3-6). Spanning the entire amino acid, there are 53 KLK3 specific amino acids and 6 KLK2 specific amino acids when comparing the gorilla chimeric KLK to the ancestral KLK3 and KLK2 amino acid chains.

### 3.3.5.2 Protein Stability

Though I have a predicted amino acid sequence, the combination of exons may not have yielded a stable protein structure. I used ExPASy's ProtParam to determine the stability of the proposed amino acid. Most notably, if expressed in mammalian reticulocytes, the gorilla chimeric KLK protein is expected to have a 30 hour half-life, with an instability index of 37.15. I then compared this to the instability rating in either human KLK3 at 39.88 (stable) or human KLK2 at 46.03 (unstable), both with half-lives greater than 30 hours in

mammalian reticulocytes. This suggest that the gorilla chimeric KLK is as stable, if not more stable than its progenitors.

## 3.4 Discussion

### 3.4.1 Successful determination of chimeric gorilla *KLK*

The structure of the chimeric gorilla *KLK* region was, by and large, determined through my ability to successfully amplify and sequence the 10kb region starting upstream of the first exon of what we are now calling the gorilla chimeric *KLK* and extending through the centromeric end of the *KLKP1* pseudogene. This amplification provided me with a single fragment that excluded regional homologous *KLK* sequence, giving me the ability to perform stepwise sequencing analysis across the entire fragment.

Importantly, the 10kb long range fragment gave me the ability to sequence across three gaps that existed in the UCSC public assembly. The two smaller gaps were sequenced which included the 271bp gap in the second intron, and the 230bp gap in the 3' UTR of the chimeric gorilla *KLK*. This last gap represents a small deletion, decreasing the expected size of the 3' UTR.

The third gap of 1105bp was the most important to understanding the chimeric gorilla *KLK*. This is the region within which the large deletion between the fourth introns of the ancestral *KLK3* and *KLK2* occurred. Once I had completely sequenced through this region, I was able to compare it to the homologous regions in both human *KLK3* and *KLK2*, and hence determine the most likely area within which the break occurred.

The *Gorilla gorilla* trace archives strengthened my results. Even though the trace archives alone do not provide sufficient coverage across chimeric gorilla *KLK* locus, the sequences from the trace archives allowed me to increase the coverage across the sequence

produced from the 10kb fragment. Added to this, the trace archives also allowed me to expand the number of individuals used to infer the genomic sequence of this area in gorilla.

However, no BACs were detected that conclusively spanned the region, and the UCSC *Gorilla gorilla* public assembly was unreliable. There are many gaps in the assembly that are difficult to reconcile. Importantly, the region in which the deletion occurred is not available on the public assembly. The reason for this is that no fragment used for the public assembly spans that region. The gap in which the chimeric deletion occurred coincides with the lack of information within the assembly. The other gaps in the public assembly coincide with repetitive DNA. This lack of information, either in the public assembly, or within the BAC library, led to the importance of the in-house long range amplification of the region.

### 3.4.2 Gorillas (and gibbons) use one protein where other catarrhines use two.

With respect to gorilla, the deletion event yielded a gene in which the first 4 exons code for (excluding the Arg<sup>45</sup>His) a completely conserved KLK3 polypeptide, and the last exon codes for a completely conserved KLK2 polypeptide. The breakpoint, and the Arg<sup>45</sup>His, occurred in the same regions of both species. When gene duplication occurs, there are multiple fates for the duplicated genes, including pseudogenization, conservation of gene function, subfunctionalization, neofunctionalization (Zhang 2003). Similar fates can be considered when thinking about the chimeric gorilla KLK. With this in mind, I will focus on the possible function of the chimeric gorilla KLK through the light of the evolutionary fate of new genes, using evidence from domain/site mapping.

There are a few different ways to think about the function of the chimeric gorilla and the similarly chimeric gibbon. The first hypothesis is that the deletion event resulted in a chimeric gorilla *KLK* pseudogene. There are examples of the *KLK2* being pseudogenized in or deleted from the genomes of multiple species, but these examples only occur in species unaffected by the duplication event that had created the *KLK3/KLK2* homolog in catarrhines. For example, although the mouse and rat ancestor only had *KLK2*, it served similar functions associated with seminal dissolution. The mouse and rat only have a *KLK2* pseudogene within the locus (Pavlopoulou *et al.* 2010). However, some elements should be considered in relation to this hypothesis. First, the entire coding region of the chimeric gorilla *KLK* is conserved for either or both *KLK3* and *KLK2* codons. Second, though mouse and rat have a *KLK2* pseudogene, but they also have multiple *KLK1* duplications that may have changed the adaptive terrain in those species (Pavlopoulou *et al.* 2010).

Another hypothesis is that only a *KLK3*-like gene function was conserved after the deletion event. Like the Arg<sup>45</sup>His, there are 6 total sites which are *KLK2*-like, but none of them are within the catalytic motifs. In gorilla, there are 53 amino acids that are *KLK3*-like, 6 of which fall in the second and third catalytic motifs. Therefore, of the 45 amino acids that fall within the catalytic motifs, 39 are shared and 6 are *KLK3*-like. This suggests that, at least for the chimeric gorilla *KLK*, the *KLK3*-like catalytic behavior is conserved. Also, one of the *KLK3*-like amino acids in the third catalytic motif, Ser<sup>206</sup>, coincides with a substrate binding site. This would indicate that the *KLK3*-like catalytic behavior that is conserved is specifically related to substrate binding.

The next hypothesis naturally follows the last, in that it could be hypothesized that only a KLK2-like gene function was conserved after the deletion event. As discussed in the previous hypothesis, the majority of differences, and their placement do not favor a KLK2-like catalytic function. Regardless, there are some points which support this hypothesis. Primarily, in spite of there being 6 KLK3-like amino acid changes in the second and third catalytic motifs, 39 are still shared as both KLK3 and KLK2-like. Additionally, all 3 amino acids that make up the catalytic triad (active sites) for both KLK3 and KLK2 are the same, and are conserved. Also, the non-synonymous amino acid change of Arg<sup>45</sup>His should be considered. This change occurred in the second exon, derived from the KLK3 ancestor, but represents a KLK2-like change.

Another hypothesis is the neofunctionalization and/or reverse-subfunctionalization of the chimeric gorilla KLK. In accordance with this hypothesis, the combination of both KLK3-like and KLK2-like properties of the polypeptide may have resulted in a function related to this combination. Further, loss of conservation of subfunction for one or both of the KLKs may have allowed selection or drift to favor the chimeric gorilla KLK, leading to a gene that is similar to the progenitor. It is known that gene duplication events leading to neofunctionalization result in an increase in nonsynonymous changes within that gene (Zhang *et al.* 1998). However, when neofunctionalization leads to a related function (Zhang 2003), and since the function of the progenitor KLK may be conserved, this would allow neofunctionalization with little to no nonsynonymous change outside of the chimeric combination of exons. The factors indicated in the previous two hypotheses also support this last one and, unlike the other two, they do not contradict each other in this context.

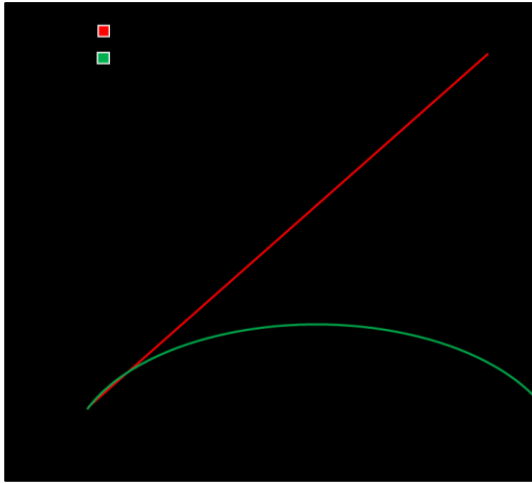


However, when considering this hypothesis, it is important to keep in mind that neofunctionalization may represent both a new function in the genome of gorilla, as well as conservation of ancestral function.

### 3.4.3 Relationship to mating system

Although both gibbon and gorilla have very different mating systems and associated physiologies, they do have important features in common: they are both monandrous, with a characteristic absence in sperm competition. . Gorillas and gibbons both have small testes relative to body size, and reduced seminal viscosity.

In apes, as sperm competition increases, so does the viscosity of the seminal fluid (Dixson & Anderson 2002). Seminal viscosity is associated with the semenogelins (SEMG1, SEMG2) and their proposed covalent cross-linking by transglutaminase (TGM4). Conversely, the proteolytic activity of KLK3, KLK2, and ACPP (as discussed earlier) cause seminal liquefaction (Lovgren *et al.* 1997). Seminal liquefaction releases the sperm from the ejaculate, allowing it to pass into cervical mucosa. In humans, a species with moderate to low sperm competition, there is a fairly rapid liquefaction of a moderately coagulated seminal fluid. Conversely, in chimpanzee there is limited seminal liquefaction. This is likely caused by a reduction of seminal proteases and a large increase in the number and density of covalently cross linked semenogelins. In gorilla, this is associated with a less viscous ejaculate and the pseudogenization of SEMG1, SEMG2, and TGM4 (Jensen-Seaman & Li 2003; Clark & Swanson 2005). In spite of this, it may be necessary to maintain some viscosity outside of sperm competition and some liquefaction within sperm competition. However, this activity should be expected to be greatly reduced (Figure 3-7).



**Figure 3-7 Proteolysis is reduced in low and high sperm competition conditions.**

Since gene duplication can lead to a new function, in the case of KLK3 and KLK2 the functions may be new but they also remain related. This may indicate that initially, if not to some extent continuously, the duplication may have been selected for as a way to favor increased gene product. Since chimpanzee has a high level of sperm competition, the rate of liquefaction decreases but the substrate for KLK3 and KLK2 remains, as may some necessity for limited proteolysis. Similarly, the low levels of coagulation in gorilla and gibbon are related to a reduction and loss of substrate (SEMG1 and SEMG2). If the initial duplication of KLK3 and KLK2 was in response to selecting for greater amounts of gene product, and the genes themselves did not diverge greatly, it is possible that this reduction in substrate is related to the loss of conservation for both KLKs.

It is likely that behavior and gene loss were related, as both gorillas and gibbons are monandrous. If this gene loss was adaptive for monandry but deleterious for polyandry, it

may limit the future reproductive strategies of both gibbon and gorilla, since genes can not be easily regained once they are lost. However, given the structure of the *KLK* genomic region, with its tandemly arrayed paralogs, regain of a gene via unequal crossing over is at least relatively easier than re-evolving a single copy gene such as *TGM4*, also lost in gorilla and gibbon.

#### 3.4.4 Adding to Published Results

The hypothesis first formed by Clark and Swanson (2005) that *KLK2* is lost or a pseudogene in gorilla and gibbon, was due, in part, to the inability to amplify exons 2, 3, and 4 of *KLK2*. This hypothesis was verified by my analysis, as I was able to describe the loss of the last exon of *KLK3* all the way through the fourth exon of *KLK2* in gorilla. Additionally, Clark and Swanson (2005) furthered the analysis by showing that a frameshift mutation occurs in the coding region of *TGM4* in both gibbon and gorilla. This is of importance due to the connection between the *KLK*'s and *TGM4*. Whereas the seminal *KLK*'s are associated with liquefaction, *TGM4* is associated with coagulation. If there is a loss in seminal coagulation, and the gene production associated with this trait, there may be no force conserving the gene associated with liquefaction.

During the course of this project, Marques et al. (2012) published similar results. They noted a deletion between the fourth exon of *KLK3* and the fifth exon of *KLK2* in both gorilla and gibbon within the same LINE2 element. They found that the first 4 exons coded for the ancestral *KLK3* amino acids, and the last exon coded for the ancestral *KLK2* amino acids. Notably, they also concluded that these deletions yielded single chimeric *KLK* genes. They

also noted five of the six amino acid that were specific to KLK2 (they did not include Arg<sup>45</sup>His), concluding that the protein maintained its KLK3-like function. Interestingly, they noted complete loss of KLK3 and KLK2 in *C. guereza*, as well as multiple loss of function mutations in several primate species. Though they did not find or note the Arg<sup>45</sup>His in the chimeric gorilla and gibbon KLK, they did note that the site (site 45) as well as 4 other sites in KLK3 and 8 sites in KLK2 were under positive selection.

Notably, the work of Jensen-Seaman and Li (2003) showed a correlation between physiological differences in seminal viscosity between species were associated with structural changes in the SEMG1 and SEMG2 of different species. They hypothesized that these differences were associated with mating system. Marques et al. (2012) further this correlation between the number of SEMG1 and SEMG2 repeat units by including the presence and activity of KLK3 and KLK2 into the correlation between seminal phenotype and mating system.

The most notable difference between the analysis of Marques et al. (2012) and my own refers to the recognition of an Arg<sup>45</sup>His in the chimeric gorilla and gibbon KLK. This change exists in both my 10kb fragment of the gorilla sequence and within the trace archives. This difference also exists in the UCSC public assembly of gorilla and gibbon, though as I noted earlier, these assemblies are still not complete in these regions and are not as useful as other assemblies. Along with that caution, it should be noted that all KLK3-containing primates with public assemblies have Arg<sup>45</sup> at the site predicted to be under positive selection by Marques et al. (2012). My interpretation of Arg<sup>45</sup>His is that this site was under positive selection in or leading to chimeric gorilla and gibbon KLK.

## 4 References

- Agrawal A.F. (2001) Sexual selection and the maintenance of sexual reproduction. *Nature* **411**, 692-5.
- Anderson M.J., Chapman S.J., Videan E.N., Evans E., Fritz J., Stoinski T.S., Dixson A.F. & Gagneux P. (2007) Functional evidence for differences in sperm competition in humans and chimpanzees. *Am J Phys Anthropol* **134**, 274-80.
- Anderson M.J. & Dixson A.F. (2002) Sperm competition: motility and the midpiece in primates. *Nature* **416**, 496.
- Andersson M.B. (1994) *Sexual selection*. Princeton University Press, Princeton, N.J.
- Anton S.C., Potts R. & Aiello L.C. (2014) Human evolution. Evolution of early *Homo*: an integrated biological perspective. *Science* **345**, 1236828.
- Arnold M.I. & Davidson E.H. (1997) The hardwiring of development: organization and function of genomic regulatory systems. *Development* **124**, 1851-64.
- Atsalis S. & Videan E. (2009) Reproductive aging in captive and wild common chimpanzees: factors influencing the rate of follicular depletion. *Am J Primatol* **71**, 271-82.
- Baker R. (1996) *Sperm wars*. BasicBooks. New York.
- Banas B., Blaschke D., Fittler F. & Horz W. (1994) Analysis of the promoter of the human prostatic acid phosphatase gene. *Biochim Biophys Acta* **1217**, 188-94.
- Bartlett T. (2003) Intragroup and intergroup social interactions in white-handed gibbons. *Int J Primatol* **24**, 239-59.
- Bartlett T.Q. (2007) The hylobatidae: small apes of Asia. *Primates in perspective*, 274-89.
- Bermejo M. (1997) Study of western lowland gorillas in the Lossi Forest of North Congo and a pilot gorilla tourism plan. *Gorilla Conservation News* **11**, 6-7.
- Birkhead T.R. & Møller A.P. (1998) *Sperm competition and sexual selection*. Academic Press., Waltham.
- Boesch C. (1996) Social grouping in Tai'chimpanzees. In: *Great ape societies*. (ed. by McGrew W.C., Marchant L.F., Nishida T.) pp. 101-13. Cambridge University Press, Cambridge, England.
- Borgono C.A., Michael I.P. & Diamandis E.P. (2004) Human tissue kallikreins: physiologic roles and applications in cancer. *Mol Cancer Res* **2**, 257-80.
- Borgoño C.A., Michael I.P., Komatsu N., Jayakumar A., Kapadia R., Clayman G.L., Sotiropoulou G. & Diamandis E.P. (2007) A potential role for multiple tissue kallikrein serine proteases in epidermal desquamation. *J Biol Chem* **282**, 3640-52.
- Bradley B.J., Robbins M.M., Williamson E.A., Steklis H.D., Steklis N.G., Eckhardt N., Boesch C. & Vigilant L. (2005) Mountain gorilla tug-of-war: Silverbacks have limited control over reproduction in multimale groups. *Proc Natl Acad Sci U S A* **102**, 9418-23.
- Brattsand M., Stefansson K., Lundh C., Haasum Y. & Egelrud T. (2005) A proteolytic cascade of kallikreins in the stratum corneum. *J Invest Dermatol* **124**, 198-203.
- Breuer T. (2008) Male reproductive success in wild western gorillas (*Gorilla gorilla*). PhD thesis, University of Leipzig, Leipzig.
- Brillard-Bourdet M., Rehault S., Juliano L., Ferrer M., Moreau T. & Gauthier F. (2002) Amidolytic activity of prostatic acid phosphatase on human semenogelins and semenogelin-derived synthetic substrates. *Eur J Biochem* **269**, 390-5.

- Brockelman W.Y., Reichard U., Treesucon U. & Raemaekers J.J. (1998) Dispersal, pair formation and social structure in gibbons (*Hylobates lar*). *Behav Ecol and Sociobiology* **42**, 329-39.
- Brown D.E. (1991) *Human universals*. McGraw-Hill New York.
- Carbone L., Alan Harris R., Gnerre S., Veeramah K.R., Lorente-Galdos B., Huddleston J., Meyer T.J., Herrero J., Roos C., Aken B., *et al.* (2014) Gibbon genome and the fast karyotype evolution of small apes. *Nature* **513**, 195-201.
- Carbone L., Harris R.A., Vessere G.M., Mootnick A.R., Humphray S., Rogers J., Kim S.K., Wall J.D., Martin D., Jurka J., Milosavljevic A. & de Jong P.J. (2009) Evolutionary breakpoints in the gibbon suggest association between cytosine methylation and karyotype evolution. *PLoS Genet* **5**, e1000538.
- Carnahan-Craig S.J. & Jensen-Seaman M.I. (2014) Rates of evolution of hominoid seminal proteins are correlated with function and expression, rather than mating system. *J Mol Evol* **78**, 87-99.
- Carnahan-Craig S.J. (2013) Molecular evolution of hominoid seminal plasma genes. PhD thesis, Duquesne University, Pittsburgh, PA.
- Carnahan S.J. & Jensen-Seaman M.I. (2008) Hominoid seminal protein evolution and ancestral mating behavior. *Am J Primatol* **70**, 939-48.
- Caro T. (2005) *Antipredator defenses in birds and mammals*. University of Chicago Press, Chicago.
- Carpenter C.R. (1984) *A field study in Siam of the behavior and social relations of the gibbon (Hylobates lar)*. AMS Press, New York, N.Y.
- Chapais B. (2009) *Primeval kinship: How pair-bonding gave birth to human society*. Harvard University Press, Cambridge.
- Chapais B. (2011a) Anthropology. The deep social structure of humankind. *Science* **331**, 1276-7.
- Chapais B. (2011b) The evolutionary history of pair-bonding and parental collaboration. *The Oxford handbook of evolutionary family psychology*, 33-50. Oxford University Press, New York.
- Chapais B. (2013) Monogamy, strongly bonded groups, and the evolution of human social structure. *Evol Anthropol* **22**, 52-65.
- Chapman T., Liddle L.F., Kalb J.M., Wolfner M.F. & Partridge L. (1995) Cost of mating in *Drosophila melanogaster* females is mediated by male accessory gland products. *Nature* **373**, 241-4.
- Chivers D.J. (1980) *Malayan forest primates : ten years' study in tropical rain forest*. Plenum Press, New York.
- Chuang T.D., Chen S.J., Lin F.F., Veeramani S., Kumar S., Batra S.K., Tu Y. & Lin M.F. (2010) Human prostatic acid phosphatase, an authentic tyrosine phosphatase, dephosphorylates ErbB-2 and regulates prostate cancer cell growth. *J Biol Chem* **285**, 23598-606.
- Clark N.L. & Swanson W.J. (2005) Pervasive adaptive evolution in primate seminal proteins. *PLoS Genet* **1**, e35.
- Clemens Z., Merker B. & Ujhelyi M. (2008) Observations on paternal care in a captive family of white-handed gibbons (*Hylobates lar*). *Gibbon Journal*, 46.
- Clutton-Brock T. & Harvey P. (1979) Comparison and adaptation. *Proc R Soc Lond [Biol]* **205**, 547-65.

- Clutton-Brock T.H. & Harvey P.H. (1977) Primate ecology and social organization. *J Zool* **183**, 1-39.
- Coller H.A. & Kruglyak L. (2008) It's the sequence, stupid! *Science* **322**, 380-1.
- Costa P.T., Jr., Terracciano A. & McCrae R.R. (2001) Gender differences in personality traits across cultures: robust and surprising findings. *J Pers Soc Psychol* **81**, 322-31.
- Courage A., Henderson I. & Watkin J. (2001) Gorilla Molecular Ecology: From the Forest to the Lab. *Gorilla Journal*, 31-5.
- Czekala N.M. & Robbins M.M. (2001) Assessment of reproduction and stress through hormone analysis in gorillas. In: *Mountain Gorillas: Three Decades of Research at Karisoke* (eds. by Robbins MM, Sicotte P & Stewart KJ), pp. 317 - 39. Cambridge University Press, Cambridge.
- Dahl J.F. & Nadler R.D. (1992) Genital swelling in females of the monogamous gibbon, *Hylobates (H.) lar*. *Am J Phys Anthropol* **89**, 101-8.
- Darwin C. (1871) *The descent of man, and selection in relation to sex*. D. Appleton and company, New York.
- Das R., Hergenrother S.D., Soto-Calderón I.D., Dew J.L., Anthony N.M. & Jensen-Seaman M.I. (2014) Complete mitochondrial genome sequence of the eastern gorilla (*Gorilla beringei*) and Implications for African Ape Biogeography. *J Hered.* **6**, 752-61.
- Dave B.N. & Rindani T.H. (1988) Acid phosphatase activity in human semen. *Int J Fertil* **33**, 45-7.
- Debela M., Beaufort N., Magdolen V., Schechter N.M., Craik C.S., Schmitt M., Bode W. & Goettig P. (2008) Structures and specificity of the human kallikrein-related peptidases KLK 4, 5, 6, and 7. *Biol Chem* **389**, 623-32.
- Descargues P., Deraison C., Bonnart C., Kreft M., Kishibe M., Ishida-Yamamoto A., Elias P., Barrandon Y., Zambruno G., Sonnenberg A. & Hovnanian A. (2005) Spink5-deficient mice mimic Netherton syndrome through degradation of desmoglein 1 by epidermal protease hyperactivity. *Nat Genet* **37**, 56-65.
- Diamandis E.P., Yousef G.M., Luo L.-Y., Magklara A. & Obiezu C.V. (2000) The new human kallikrein gene family: implications in carcinogenesis. *TEM* **11**, 54-60.
- Dixon A.F. (1995) Sexual selection and ejaculatory frequencies in primates. *Folia Primatol (Basel)* **64**, 146-52.
- Dixon A.F. (2009) *Sexual selection and the origins of human mating systems*. Oxford University Press, Oxford, New York.
- Dixon A.F. (2012) *Primate sexuality : comparative studies of the prosimians, monkeys, apes, and humans*. Oxford University Press, Oxford.
- Dixon A.L. & Anderson M.J. (2002) Sexual selection, seminal coagulation and copulatory plug formation in primates. *Folia Primatol (Basel)* **73**, 63-9.
- Doran D.M. (1996) Comparative positional behavior of the African apes. In: *Great ape societies*. (ed. by McGrew W.C., Marchant L.F., Nishida T.) pp. 213-24. Cambridge University Press, Cambridge, England.
- Dorus S., Evans P.D., Wyckoff G.J., Choi S.S. & Lahn B.T. (2004) Rate of molecular evolution of the seminal protein gene *SEMG2* correlates with levels of female promiscuity. *Nat Genet* **36**, 1326-9.
- Felsenstein J. (1974) The evolutionary advantage of recombination. *Genetics* **78**, 737-56.

- Aureli F., Schaffner C.M., Boesch C., Bearder S.K., Call J., Chapman C.A., Connor R., Di Fiore A., Dunbar R., Henzi S., *et al.* (2008) Fission-fusion dynamics: new research frameworks. *Curr Anthropol* **49**, 627-54.
- Fleagle J.G. (1999) *Primate adaptation and evolution*. Academic Press, San Diego.
- Fletcher A. & Fletcher A. (2001) *Development of infant independence from the mother in wild mountain gorillas*. *Mountain Gorillas*. Cambridge University Press.
- Fossey D. & Harcourt A.H. (1977) Feeding ecology of the free ranging mountain gorilla (*Gorilla gorilla beringei*). In: *Primate ecology : studies of feeding and ranging behaviour in lemurs, monkeys, and apes* (ed. by Clutton-Brock TH), pp. 415-447. Academic Press, London ; New York.
- Fowler K. & Partridge L. (1989) A cost of mating in female fruitflies. *Nature* **338**, 760-1.
- Furuichi T. (2011) Female contributions to the peaceful nature of bonobo society. *Evol Anthro* **20**, 131-42.
- Gagneux P., Boesch C. & Woodruff D.S. (1999) Female reproductive strategies, paternity and community structure in wild West African chimpanzees. *Anim Behav* **57**, 19-32.
- Galdikas B.M. & Wood J.W. (1990) Birth spacing patterns in humans and apes. *Am J Phys Anthropol* **83**, 185-91.
- Galdikas B.M.F. (1995) Social and Reproductive Behavior of Wild Adolescent Female Orangutans. In: *The Neglected Ape* (eds. by Nadler R, Galdikas BM, Sheeran L & Rosen N), pp. 163-82. Springer US.
- Gavrilets S. (2012) Human origins and the transition from promiscuity to pair-bonding. *Proc Natl Acad Sci U S A* **109**, 9923-8.
- Geary D.C. (2004) Sexual selection and reproductive competition in primates: New perspectives and directions. *American Journal of Human Biology* **16**, 176-8.
- Geissmann T. (1991) Reassessment of age of sexual maturity in gibbons (*Hylobates* spp.). *Am J Primatol* **23**, 11-22.
- Geissmann T., Geschke K. & Blanchard B. (2009) Longevity in gibbons (Hylobatidae). *Gibbon Journal* **5**, 81-92.
- Goldberg T.L. & Wrangham R.W. (1997) Genetic correlates of social behaviour in wild chimpanzees: evidence from mitochondrial DNA. *Anim Behav* **54**, 559-70.
- Goodall J. (1986) *The chimpanzees of Gombe: patterns of behavior*. Belknap Press, London.
- Goodman M., Porter C.A., Czelusniak J., Page S.L., Schneider H., Shoshani J., Gunnell G. & Groves C.P. (1998) Toward a phylogenetic classification of Primates based on DNA evidence complemented by fossil evidence. *Mol Phylogenet Evol* **9**, 585-98.
- Graddis T.J., McMahan C.J., Tamman J., Page K.J. & Trager J.B. (2011) Prostatic acid phosphatase expression in human tissues. *Int J Clin Exp Pathol* **4**, 295-306.
- Green R.E., Krause J., Briggs A.W., Maricic T., Stenzel U., Kircher M., Patterson N., Li H., Zhai W., Fritz M.H., *et al.* (2010) A draft sequence of the Neandertal genome. *Science* **328**, 710-22.
- Groves C.P. (1970) Population systematics of the gorilla. *J Zool* **161**, 287-300.
- Groves C.P. (2001) *Primate taxonomy*. Smithsonian Institution Press, Washington, DC.
- Gruber T., Clay Z. & Zuberbühler K. (2010) A comparison of bonobo and chimpanzee tool use: evidence for a female bias in the *Pan* lineage. *Anim Behav* **80**, 1023-33.
- Haeckel E. (1873) *Natürliche Schöpfungsgeschichte: I*. Kröner.



- Haimoff E.H., Chivers D.J., Gittins S.P. & Whitten T. (1982) A phylogeny of gibbons (*Hylobates spp.*) based on morphological and behavioural characters. *Folia Primatol (Basel)* **39**, 213-37.
- Harcourt A., Purvis A. & Liles L. (1995) Sperm competition: mating system, not breeding season, affects testes size of primates. *Funct Ecol*, **9**, 468-76.
- Harcourt A. & Stewart K. (2007) Gorilla society: conflict, compromise, and cooperation between the sexes. *mammalia*, **72**, 139-147.
- Harcourt A.H. (1978) Strategies of emigration and transfer by primates, with particular reference to gorillas. *Z Tierpsychol* **48**, 401-20.
- Harcourt A.H., Harvey P.H., Larson S.G. & Short R. (1981) Testis weight, body weight and breeding system in primates. *Nature* **293**, 55-7.
- Harding, R. S., & Teleki, G. (1981). *Omnivorous primates: gathering and hunting in human evolution*. Columbia University Press, New York.
- Hardus M., Lameira A., Zulfa A., Atmoko S.S., de Vries H. & Wich S. (2012) Behavioral, Ecological, and Evolutionary Aspects of Meat-Eating by Sumatran Orangutans (*Pongo abelii*). *Int J Primatol* **33**, 287-304.
- Harrison T. (2010) Apes among the tangled branches of human origins. *Science* **327**, 532-4.
- Hartwig W.C. (2002) *The primate fossil record*. Cambridge University Press, Cambridge ; New York.
- Henn B.M., Cavalli-Sforza L.L. & Feldman M.W. (2012) The great human expansion. *Proc Natl Acad Sci U S A* **109**, 17758-64.
- Holden C. & Mace R. (1997) Phylogenetic analysis of the evolution of lactose digestion in adults. *Hum Biol* **69**, 605-28.
- Holland B. & Rice W.R. (1998) Perspective: chase-away sexual selection: antagonistic seduction versus resistance. *Evolution*, **52**, 1-7.
- Hrdy S.B. (2000) The optimal number of fathers: evolution, demography, and history in the shaping of female mate preferences. *Ann NY Acad Sci* **907**, 75-96.
- Hrdy S.B. (2009) *Mothers and others : the evolutionary origins of mutual understanding*. Belknap Press of Harvard University Press, Cambridge, Mass.
- Insler V., Glezerman M., Zeidel L., Bernstein D. & Misgav N. (1980) Sperm storage in the human cervix: a quantitative study. *Fertil Steril* **33**, 288-93.
- Jakob C.G., Lewinski K., Kuciel R., Ostrowski W. & Lebiada L. (2000) Crystal structure of human prostatic acid phosphatase. *The Prostate* **42**, 211-8.
- Jameson N.M., Hou Z.C., Sterner K.N., Weckle A., Goodman M., Steiper M.E. & Wildman D.E. (2011) Genomic data reject the hypothesis of a prosimian primate clade. *J Hum Evol* **61**, 295-305.
- Jensen-Seaman M.I., Deinard A.S. & Kidd K.K. (2001) Modern African ape populations as genetic and demographic models of the last common ancestor of humans, chimpanzees, and gorillas. *J Hered* **92**, 475-80.
- Jensen-Seaman M.I. & Li W.H. (2003) Evolution of the hominoid semenogelin genes, the major proteins of ejaculated semen. *J Mol Evol* **57**, 261-70.
- Jurka J. (2000) Repbase update: a database and an electronic journal of repetitive elements. *Trends Genet* **16**, 418-20.
- Kawasaki K., Hu J.C. & Simmer J.P. (2014) Evolution of KLK4 and enamel maturation in eutherians. *Biol Chem* **395**, 1003-13.

- Kent W.J., Sugnet C.W., Furey T.S., Roskin K.M., Pringle T.H., Zahler A.M. & Haussler D. (2002) The human genome browser at UCSC. *Genome Res* **12**, 996-1006.
- Kimura M. (1979) The neutral theory of molecular evolution. *Sci Am* **241**, 98-100.
- King M.C. & Wilson A.C. (1975) Evolution at two levels in humans and chimpanzees. *Science* **188**, 107-16.
- Knott C.D., Emery Thompson M., Stumpf R.M. & McIntyre M.H. (2010) Female reproductive strategies in orangutans, evidence for female choice and counterstrategies to infanticide in a species with frequent sexual coercion. *Proc R Soc Lond [Biol]* **277**, 105-13.
- Kong H.Y. & Byun J. (2013) Emerging roles of human prostatic Acid phosphatase. *Biomol Ther (Seoul)* **21**, 10-20.
- Koumandou V.L. & Scorilas A. (2013) Evolution of the plasma and tissue kallikreins, and their alternative splicing isoforms. *PLoS One* **8**, e68074.
- Kraut H., Frey E.K. & Werle E. (1930) Der Nachweis eines Kreislaufhormons in der Pankreasdrüse. (IV. Mitteilung über dieses Kreislaufhormon.). In: *Hoppe-Seyler's Zeitschrift für physiologische Chemie*, p. 97.
- Krem M.M. & Di Cera E. (2001) Molecular markers of serine protease evolution. *EMBO J* **20**, 3036-45.
- Kuciel R., Bakalova A., Mazurkiewicz A., Bilska A. & Ostrowski W. (1990) Is the subunit of prostatic phosphatase active? Reversible denaturation of prostatic acid phosphatase. *Biochem Int* **22**, 329-34.
- Lad P.M., Learn D.B., Cooper J.F. & Reisinger D.M. (1984) Distribution of prostatic acid phosphatase isoenzymes in normal and cancerous states. *Clin Chim Acta* **141**, 51-65.
- Lappan S. (2007) Social Relationships among Males in Multimale Siamang Groups. *Int J Primatol* **28**, 369-87.
- Lawrence M.G., Lai J. & Clements J.A. (2010) Kallikreins on steroids: structure, function, and hormonal regulation of prostate-specific antigen and the extended kallikrein locus. *Endocr Rev* **31**, 407-46.
- Lee C.L., Li S.S. & Chu T.M. (1984) Immunologically reactive tryptic fragments of human prostatic acid phosphatase. *Biochem J* **223**, 871-7.
- Lin M.F. & Garcia-Arenas R. (1994) Effect of cell density on androgen regulation of the mRNA level of human prostatic acid phosphatase. *Mol Cell Endocrinol* **99**, R21-4.
- Locke D.P., Hillier L.W., Warren W.C., Worley K.C., Nazareth L.V., Muzny D.M., Yang S.P., Wang Z., Chinwalla A.T., Minx P., *et al.* (2011) Comparative and demographic analysis of orang-utan genomes. *Nature* **469**, 529-33.
- Loeb S. & Catalona W.J. (2007) Prostate-specific antigen in clinical practice. *Cancer Lett* **249**, 30-9.
- Lovejoy C.O. (2009) Reexamining human origins in light of *Ardipithecus ramidus*. *Science* **326**, 74e1-8.
- Lovgren J., Rajakoski K., Karp M., Lundwall A. & Lilja H. (1997) Activation of the zymogen form of prostate-specific antigen by human glandular kallikrein 2. *Biochem Biophys Res Commun* **238**, 549-55.
- Lundwall A. (2013) Old genes and new genes: the evolution of the kallikrein locus. *Thromb Haemost* **110**, 469-75.
- Lundwall A. & Brattsand M. (2008) Kallikrein-related peptidases. *Cell Mol Life Sci* **65**, 2019-38.

- Luo L., Herbrick J.-A., Scherer S.W., Beatty B., Squire J. & Diamandis E.P. (1998) Structural characterization and mapping of the normal epithelial cell-specific 1 gene. *Biochem Biophys Res Commun* **247**, 580-6.
- Maestripieri D. & Roney J.R. (2005) Primate copulation calls and postcopulatory female choice. *Behav Ecol* **16**, 106-13.
- Maggioncalda A.N., Czekala N.M. & Sapolsky R.M. (2002) Male orangutan subadulthood: a new twist on the relationship between chronic stress and developmental arrest. *Am J Phys Anthropol* **118**, 25-32.
- Mailund T., Dutheil J.Y., Hobolth A., Lunter G. & Schierup M.H. (2011) Estimating divergence time and ancestral effective population size of Bornean and Sumatran orangutan subspecies using a coalescent hidden Markov model. *PLoS Genet* **7**, e1001319.
- Markham R. & Groves C.P. (1990) Brief communication: Weights of wild orang utans. *Am J Phys Anthropol* **81**, 1-3.
- Marques P.I., Bernardino R., Fernandes T., Green E.D., Hurle B., Quesada V. & Seixas S. (2012) Birth-and-death of *KLK3* and *KLK2* in primates: evolution driven by reproductive biology. *Genome Biol Evol* **4**, 1331-8.
- McDougall I., Brown F.H. & Fleagle J.G. (2005) Stratigraphic placement and age of modern humans from Kibish, Ethiopia. *Nature* **433**, 733-6.
- McLean C.Y., Reno P.L., Pollen A.A., Bassan A.I., Capellini T.D., Guenther C., Indjeian V.B., Lim X., Menke D.B., Schaar B.T., Wenger A.M., Bejerano G. & Kingsley D.M. (2011) Human-specific loss of regulatory DNA and the evolution of human-specific traits. *Nature* **471**, 216-9.
- Meijaard E., Wich S., Ancrenaz M. & Marshall A.J. (2012) Not by science alone: why orangutan conservationists must think outside the box. *Ann NY Acad Sci* **1249**, 29-44.
- Meredith R.W., Gatesy J., Murphy W.J., Ryder O.A. & Springer M.S. (2009) Molecular decay of the tooth gene Enamelin (ENAM) mirrors the loss of enamel in the fossil record of placental mammals. *PLoS Genet* **5**, e1000634.
- Mitchell G., Erwin J. & Swindler D.R. (1986) *Comparative primate biology*. A.R. Liss, New York.
- Moore H.D., Martin M. & Birkhead T.R. (1999) No evidence for killer sperm or other selective interactions between human spermatozoa in ejaculates of different males *in vitro*. *Proc Biol Sci* **266**, 2343-50.
- Mootnick A. & Groves C. (2005) A new generic name for the hoolock gibbon (*Hylobatidae*). *Int J Primatol* **26**, 971-6.
- Mootnick A.R. (2006) Gibbon (*Hylobatidae*) Species identification recommended for rescue or breeding centers. *Primate Conservation*, 103-38.
- Morrogh-Bernard H.C., Morf N.V., Chivers D.J. & Krützen M. (2011) Dispersal patterns of orang-utans (*Pongo spp.*) in a Bornean peat-swamp forest. *Int J Primatol* **32**, 362-76.
- Mulholland D.J., Cox M., Read J., Rennie P. & Nelson C. (2004) Androgen responsiveness of Renilla luciferase reporter vectors is promoter, transgene, and cell line dependent. *Prostate* **59**, 115-9.
- Muller H.J. (1950) Evidence of the precision of genetic adaptation. *Harvey Lect* **43**, 165-229.
- Muniyan S., Chaturvedi N.K., Dwyer J.G., Lagrange C.A., Chaney W.G. & Lin M.F. (2013) Human prostatic Acid phosphatase: structure, function and regulation. *Int J Mol Sci* **14**, 10438-64.

- Munn C. & Fernandez M. (1997) Infant development. *Orangutan species survival plan husbandry manual*. Chicago, 59-66.
- Nadler R.D. (1975) Sexual cyclicity in captive lowland gorillas. *Science* **189**, 813-4.
- Nakahashi W. & Horiuchi S. (2012) Evolution of ape and human mating systems. *J Theor Biol* **296**, 56-64.
- Napier J.R. & Napier P.H. (1967) *A handbook of living primates: morphology, ecology and behaviour of nonhuman primates*. Academic P., London, New York,.
- Newman R.W. (1970) Why man is such a sweaty and thirsty naked animal: a speculative review. *Hum Biol* **42**, 12-27.
- Nishida T. & Kawanaka K. (1985) Within-group cannibalism by adult male chimpanzees. *Primates* **26**, 274-84.
- Nowak R.M. & Walker E.P. (1999) *Walker's primates of the world*. Johns Hopkins University Press, Baltimore.
- Opie C., Atkinson Q.D. & Shultz S. (2012) The evolutionary history of primate mating systems. *Commun Integr Biol* **5**, 458-61.
- Oppenheimer S. (2012) Out-of-Africa, the peopling of continents and islands: tracing uniparental gene trees across the map. *Philos Trans R Soc London [Biol]* **367**, 770-84.
- Ortlund E., LaCount M.W. & Lebioda L. (2003) Crystal structures of human prostatic acid phosphatase in complex with a phosphate ion and alpha-benzylaminobenzylphosphonic acid update the mechanistic picture and offer new insights into inhibitor design. *Biochemistry* **42**, 383-9.
- Ostanin K., Saeed A. & Van Etten R.L. (1994) Heterologous expression of human prostatic acid phosphatase and site-directed mutagenesis of the enzyme active site. *J Biol Chem* **269**, 8971-8.
- Ozer A., Orhan F.O. & Ekerbicer H.C. (2013) Sociodemographic variables and depression in Turkish women from polygamous versus monogamous families. *Health Care Women Int* **34**, 1024-34.
- Pagel M. & Bodmer W. (2003) A naked ape would have fewer parasites. *Proc R Soc Lond [Biol]* **270**, S117-S9.
- Palombit R. (1994a) Dynamic pair bonds in *Hylobatids*: Implications regarding monogamous social systems. *Behaviour* **128**, 65-101.
- Palombit R. (1995) Longitudinal patterns of reproduction in wild female siamang (*Hylobates syndactylus*) and white-handed gibbons (*Hylobates lar*). *Int J Primatol* **16**, 739-60.
- Palombit R.A. (1994b) Extra-pair copulations in a monogamous ape. *Anim Behav* **47**, 721-3.
- Pampalakis G. & Sotiropoulou G. (2007) Tissue kallikrein proteolytic cascade pathways in normal physiology and cancer. *Biochim Biophys Acta* **1776**, 22-31.
- Patterson C. & D'Augelli A.R. (2013) *Handbook of psychology and sexual orientation*. Oxford University Press, New York.
- Pavlopoulou A., Pampalakis G., Michalopoulos I. & Sotiropoulou G. (2010) Evolutionary history of tissue kallikreins. *PLoS One* **5**, e13781.
- Pepper J.W., Mitani J.C. & Watts D.P. (1999) General gregariousness and specific social preferences among wild chimpanzees. *Int J Primatol* **20**, 613-32.
- Perry G.H., Dominy N.J., Claw K.G., Lee A.S., Fiegler H., Redon R., Werner J., Villanea F.A., Mountain J.L., Misra R., *et al.* (2007) Diet and the evolution of human amylase gene copy number variation. *Nat Genet* **39**, 1256-60.

- Peterson S.A. (1999) Marriage structure and contraception in Niger. *J Biosoc Sci* **31**, 93-104.
- Phillips D. (2014) The Prude in the Law: Why the polygamy reference is all about sex. *Appeal* **19**.
- Plavcan J.M. (2000) Inferring social behavior from sexual dimorphism in the fossil record. *J Hum Evol* **39**, 327-44.
- Plavcan J.M. (2012) Sexual size dimorphism, canine dimorphism, and male-male competition in primates: where do humans fit in? *Hum Nat* **23**, 45-67.
- Raff R.A. & Kaufman T.C. (1991) *Embryos, genes, and evolution: the developmental-genetic basis of evolutionary change*. Indiana University Press.
- Reich D., Green R.E., Kircher M., Krause J., Patterson N., Durand E.Y., Viola B., Briggs A.W., Stenzel U., Johnson P.L., *et al.* (2010) Genetic history of an archaic hominin group from Denisova Cave in Siberia. *Nature* **468**, 1053-60.
- Reichard U. (1995) Extra-pair copulations in a monogamous gibbon (*Hylobates lar*). *Ethology* **100**, 99-112.
- Reichard U. & Barelli C. (2008) Life History and Reproductive Strategies of Khao Yai *Hylobates lar*: Implications for Social Evolution in Apes. *Int J Primatol* **29**, 823-44.
- Reichard U.H. & Boesch C. (2003) *Monogamy: mating strategies and partnerships in birds, humans and other mammals*. Cambridge University Press.
- Reno P.L., Meindl R.S., McCollum M.A. & Lovejoy C.O. (2003) Sexual dimorphism in *Australopithecus afarensis* was similar to that of modern humans. *Proc Natl Acad Sci U S A* **100**, 9404-9.
- Rice W.R. & Chippindale A.K. (2001) Sexual recombination and the power of natural selection. *Science* **294**, 555-9.
- Riegman P., Vlietstra R., Klaassen P., Van der Korput J., van Kessel A.G., Romijn J. & Trapman J. (1989a) The prostate-specific antigen gene and the human glandular kallikrein-1 gene are tandemly located on chromosome 19. *FEBS Lett* **247**, 123-6.
- Riegman P., Vlietstra R., Van der Korput J., Romijn J. & Trapman J. (1989b) Characterization of the prostate-specific antigen gene: a novel human kallikrein-like gene. *Biochem Biophys Res Commun* **159**, 95-102.
- Riethoven J.J. (2010) Regulatory regions in DNA: promoters, enhancers, silencers, and insulators. *Methods Mol Biol* **674**, 33-42.
- Rijksen H.D. (1978) A field study on Sumatran orang-utans (*Pongo pygmaeus abelii* Lesson 1827): ecology, behaviour and conservation. Veenman, Wageningen.
- Robbins M.M., Bermejo M., Cipolletta C., Magliocca F., Parnell R.J. & Stokes E. (2004) Social structure and life-history patterns in western gorillas (*Gorilla gorilla gorilla*). *Am J Primatol* **64**, 145-59.
- Robbins M.M., Sicotte P. & Stewart K.J. (2005) *Mountain gorillas: three decades of research at Karisoke*. Cambridge University Press.
- Robson S.L. & Wood B. (2008) Hominin life history: reconstruction and evolution. *J Anat* **212**, 394-425.
- Rockman M.V., Hahn M.W., Soranzo N., Zimprich F., Goldstein D.B. & Wray G.A. (2005) Ancient and recent positive selection transformed opioid *cis*-regulation in humans. *PLoS Biol* **3**, e387.

- Rosenbloom K.R., Sloan C.A., Malladi V.S., Dreszer T.R., Learned K., Kirkup V.M., Wong M.C., Maddren M., Fang R., Heitner S.G., *et al.* (2013) ENCODE data in the UCSC Genome Browser: year 5 update. *Nucleic Acids Res* **41**, D56-63.
- Ross M.T., LaBrie S., McPherson J. & Stanton V.P. (2001) Screening Large-Insert Libraries by Hybridization. *Curr Protoc Hum Genet* **21**:5.6:5.6.1–5.6.52.
- Roussel J. & Austin C. (1967) Enzymic liquefaction of primate semen. *Int J Fertil* **12**, 288.
- Rowe N. (1996) *The pictorial guide to the living primates*. Pogonias Press, East Hampton, N.Y.
- Ryan C. (2011) *Sex At Dawn: How We Mate, Why We Stray, And What It Means For Modern Relationships* Harper Perennial, New York.
- Sales K.U., Masedunskas A., Bey A.L., Rasmussen A., Weigert R., List K., Szabo R., Overbeek P.A. & Bugge T.H. (2010a) Matriptase initiates epidermal prokallikrein activation and disease onset in a mouse model of Netherton syndrome. *Nat Genet* **42**, 676-83.
- Sales K.U., Masedunskas A., Bey A.L., Rasmussen A.L., Weigert R., List K., Szabo R., Overbeek P.A. & Bugge T.H. (2010b) Matriptase initiates activation of epidermal prokallikrein and disease onset in a mouse model of Netherton syndrome. *Nat Genet* **42**, 676-83.
- Sandelin A., Wasserman W.W. & Lenhard B. (2004) ConSite: web-based prediction of regulatory elements using cross-species comparison. *Nucleic Acids Res* **32**, W249-52.
- Savage D. (2011) Monogamish, [www.thestranger.com](http://www.thestranger.com) [Pop Culture].
- Schedlich L.J., Bennetts B.H. & Morris B.J. (1987) Primary structure of a human glandular kallikrein gene. *Dna* **6**, 429-37.
- Schieferstein G. (1999) Prostate-specific antigen (PSA) in human seminal plasma. *Arch Androl* **42**, 193-7.
- Schürmann C. & van Hooff J.A.R.A.M. (1986) Reproductive strategies of the orang-utan: New data and a reconsideration of existing sociosexual models. *Int J Primatol* **7**, 265-87.
- Setchell, J. M. (2003). The evolution of alternative reproductive morphs in male primates. In: *Sexual selection and reproductive competition in primates: new perspectives and direction*, pp. 413-436. American Society of Primatologists, Norman, OK.
- Shan J., Porvari K., Kivinen A., Patrikainen L., Halmekyto M., Janne J. & Vihko P. (2003) Tissue-specific expression of the prostatic acid phosphatase promoter constructs. *Biochem Biophys Res Commun* **311**, 864-9.
- Shan J., Porvari K. & Vihko P. (2005) GAAAATATGATA-like elements in androgen-associated regulation of the prostatic acid phosphatase gene. *J Steroid Biochem Mol Biol* **96**, 245-9.
- Shan J.D., Porvari K., Ruokonen M., Karhu A., Launonen V., Hedberg P., Oikarinen J. & Vihko P. (1997) Steroid-involved transcriptional regulation of human genes encoding prostatic acid phosphatase, prostate-specific antigen, and prostate-specific glandular kallikrein. *Endocrinology* **138**, 3764-70.
- Sharief F.S. & Li S.S. (1994) Nucleotide sequence of human prostatic acid phosphatase *ACPP* gene, including seven Alu repeats. *Biochem Mol Biol Int* **33**, 561-5.
- Shaw J.L. & Diamandis E.P. (2007) Distribution of 15 human kallikreins in tissues and biological fluids. *Clin Chem* **53**, 1423-32.
- Shields W. (1982) *Philopatry, inbreeding, and the evolution of sex*. SUNY Press, New York.
- Shields W. (1987) Dispersal and mating systems: investigating their causal connections. In: *Mammalian dispersal patterns* (eds. by Chepko-Sade B & Halpin Z), pp. 3-24. University of Chicago Press, Chicago.

- Shifera A.S. & Hardin J.A. (2010) Factors modulating expression of *Renilla* luciferase from control plasmids used in luciferase reporter gene assays. *Anal Biochem* **396**, 167-72.
- Shultz S., Opie C. & Atkinson Q.D. (2011) Stepwise evolution of stable sociality in primates. *Nature* **479**, 219-22.
- Shumaker R.W., Wich S.A. & Perkins L. (2008) Reproductive life history traits of female orangutans (*Pongo* spp.). *Interdiscip Top Gerontol* **36**, 147-61.
- Sicotte P. & Sicotte P. (2001) *Female mate choice in mountain gorillas*. Mountain Gorillas. Cambridge University Press.
- Singh G., Adaikan P.G. & Ng Y.K. (1996) Is seminal prostatic acid phosphatase a reliable marker for male infertility? *Singapore Med J* **37**, 598-9.
- Smuts B.B. (1987) *Primate societies*. University of Chicago Press, Chicago.
- Springer M.S., Meredith R.W., Gatesy J., Emerling C.A., Park J., Rabosky D.L., Stadler T., Steiner C., Ryder O.A., Janecka J.E., Fisher C.A. & Murphy W.J. (2012) Macroevolutionary dynamics and historical biogeography of primate diversification inferred from a species supermatrix. *PLoS One* **7**, e49521.
- Starkweather K.E. & Hames R. (2012) A survey of non-classical polyandry. *Hum Nat* **23**, 149-72.
- Steiper M.E. & Young N.M. (2006) Primate molecular divergence dates. *Mol Phylogenet Evol* **41**, 384-94.
- Stevens N.J., Seiffert E.R., O'Connor P.M., Roberts E.M., Schmitz M.D., Krause C., Gorscak E., Ngasala S., Hieronymus T.L. & Temu J. (2013) Palaeontological evidence for an Oligocene divergence between Old World monkeys and apes. *Nature* **497**, 611-4.
- Stewart K.J. & Stewart K.J. (2001) *Social relationships of immature gorillas and silverbacks*. Mountain Gorillas. Cambridge University Press.
- Stringer C. (2003) Human evolution: Out of Ethiopia. *Nature* **423**, 692-5.
- Tanaka M., Kishi Y., Takanezawa Y., Kakehi Y., Aoki J. & Arai H. (2004) Prostatic acid phosphatase degrades lysophosphatidic acid in seminal plasma. *FEBS Lett* **571**, 197-204.
- Taylor A.B. & Goldsmith M.L. (2003) *Gorilla biology : a multidisciplinary perspective*. Cambridge University Press, Cambridge.
- Thalmann O., Fischer A., Lankester F., Paabo S. & Vigilant L. (2007) The complex evolutionary history of gorillas: insights from genomic data. *Mol Biol Evol* **24**, 146-58.
- Thompson-Handler N., Malenky R.K. & Badrian N. (1984) Sexual behavior of *Pan paniscus* under natural conditions in the Lomako Forest, Equateur, Zaire. In: *The pygmy chimpanzee*, pp. 347-68. Springer, New York.
- Tomasello M. & Vaish A. (2013) Origins of human cooperation and morality. *Annu Rev Psychol* **64**, 231-55.
- United States. Central Intelligence Agency. (2014) The world factbook. Central Intelligence Agency, Washington, DC.
- Utami S.S., Goossens B., Bruford M.W., de Ruiter J.R. & van Hooff J.A.R.A.M. (2002) Male bimaturism and reproductive success in Sumatran orang-utans. *Behav Ecol* **13**, 643-52.
- Vanpé C., Kjellander P., Galan M., Cosson J.-F., Aulagnier S., Liberg O. & Hewison A.M. (2008) Mating system, sexual dimorphism, and the opportunity for sexual selection in a territorial ungulate. *Behav Ecol* **19**, 309-16.
- Verstappen H.T. (1997) The effect of climatic change on southeast Asian geomorphology. *J Quaternary Sci* **12**, 413-8.

- Vervaecke H., Van Elsacker L., Möhle U., Heistermann M. & Verheyen R.F. (1999) Inter-menstrual intervals in captive bonobos *Pan paniscus*. *Primates* **40**, 283-9.
- Vigilant L., Hofreiter M., Siedel H. & Boesch C. (2001) Paternity and relatedness in wild chimpanzee communities. *Proc Natl Acad Sci U S A* **98**, 12890-5.
- Vihko P. (1979) Human prostatic acid phosphatases: purification of a minor enzyme and comparisons of the enzymes. *Invest Urol* **16**, 349-52.
- Waal F.B.M.d. & Lanting F. (1997) *Bonobo : the forgotten ape*. University of California Press, Berkeley.
- Waltering K.K., Helenius M.A., Sahu B., Manni V., Linja M.J., Janne O.A. & Visakorpi T. (2009) Increased expression of androgen receptor sensitizes prostate cancer cells to low levels of androgens. *Cancer Res* **69**, 8141-9.
- Watts D.P. (1989) Infanticide in mountain gorillas: new cases and a reconsideration of the evidence. *Ethology* **81**, 1-18.
- White F.J. (1996) Comparative socio-ecology of *Pan paniscus*. In: *Great ape societies*. (ed. by McGrew W.C., Marchant L.F., Nishida T.) pp. 29-41. Cambridge University Press, Cambridge, England.
- White K.Y., Rodemich L., Nyalwidhe J.O., Comunale M.A., Clements M.A., Lance R.S., Schellhammer P.F., Mehta A.S., Semmes O.J. & Drake R.R. (2009) Glycomic characterization of prostate-specific antigen and prostatic acid phosphatase in prostate cancer and benign disease seminal plasma fluids. *J Proteome Res* **8**, 620-30.
- Whiten A., Goodall J., McGrew W.C., Nishida T., Reynolds V., Sugiyama Y., Tutin C.E., Wrangham R.W. & Boesch C. (1999) Cultures in chimpanzees. *Nature* **399**, 682-5.
- Wich S., Utami-Atmoko S., Setia T.M., Djoyosudharmo S. & Geurts M. (2006) Dietary and energetic responses of *Pongo abelii* to fruit availability fluctuations. *Int J Primatol* **27**, 1535-50.
- Wich S.A., Utami-Atmoko S.S., Setia T.M., Rijksen H.D., Schürmann C., van Hooft J.A.R.A.M. & van Schaik C.P. (2004) Life history of wild Sumatran orangutans (*Pongo abelii*). *J Hum Evol* **47**, 385-98.
- Wilkinson R.D., Steiper M.E., Soligo C., Martin R.D., Yang Z. & Tavaré S. (2011) Dating primate divergences through an integrated analysis of palaeontological and molecular data. *Syst Biol* **60**, 16-31.
- Wilson M.D., Barbosa-Morais N.L., Schmidt D., Conboy C.M., Vanes L., Tybulewicz V.L., Fisher E.M., Tavaré S. & Odom D.T. (2008) Species-specific transcription in mice carrying human chromosome 21. *Science* **322**, 434-8.
- Wong A. (2010) Testing the effects of mating system variation on rates of molecular evolution in primates. *Evolution* **64**, 2779-85.
- Wrangham R.W., Chapman C.A., Clark-Arcadi A.P. & Isabirye-Basuta G. (1996) Social ecology of Kanyawara chimpanzees: implications for understanding the costs of great ape groups. In: *Great ape societies*. (ed. by McGrew W.C., Marchant L.F., Nishida T.) pp. 45-57. Cambridge University Press, Cambridge, England.
- Wray G.A. (2007) The evolutionary significance of *cis*-regulatory mutations. *Nat Rev Genet* **8**, 206-16.
- Wray G.A., Hahn M.W., Abouheif E., Balhoff J.P., Pizer M., Rockman M.V. & Romano L.A. (2003) The evolution of transcriptional regulation in eukaryotes. *Mol Biol Evol* **20**, 1377-419.



- Yamagiwa J., Kahekwa J. & Basabose A. (2003) Intra-specific variation in social organization of gorillas: implications for their social evolution. *Primates* **44**, 359-69.
- Yamagiwa J., Mwanza N., Yumoto T. & Maruhashi T. (1994) Seasonal change in the composition of the diet of eastern lowland gorillas. *Primates* **35**, 1-14.
- Yousef G., Luo L. & Diamandis E. (1998) Identification of novel human kallikrein-like genes on chromosome 19q13.3-q13.4. *Anticancer Res* **19**, 2843-52.
- Yousef G.M. & Diamandis E.P. (2001) The new human tissue kallikrein gene family: structure, function, and association to disease. *Endoc Rev* **22**, 184-204.
- Yousef G.M., Luo L.Y., Scherer S.W., Sotiropoulou G. & Diamandis E.P. (1999) Molecular characterization of zyme/protease M/neurosin (PRSS9), a hormonally regulated kallikrein-like serine protease. *Genomics* **62**, 251-9.
- Yousef G.M., Scorilas A. & Diamandis E.P. (2000a) Genomic organization, mapping, tissue expression, and hormonal regulation of trypsin-like serine protease (TLSP PRSS20), a new member of the human kallikrein gene family. *Genomics* **63**, 88-96.
- Yousef G.M., Scorilas A., Magklara A., Soosaipillai A. & Diamandis E.P. (2000b) The *KLK7* (*PRSS6*) gene, encoding for the stratum corneum chymotryptic enzyme is a new member of the human kallikrein gene family—genomic characterization, mapping, tissue expression and hormonal regulation. *Gene* **254**, 119-28.
- Zelivianski S., Comeau D. & Lin M.F. (1998) Cloning and analysis of the promoter activity of the human prostatic acid phosphatase gene. *Biochem Biophys Res Commun* **245**, 108-12.
- Zelivianski S., Glowacki R. & Lin M.F. (2004) Transcriptional activation of the human prostatic acid phosphatase gene by NF-kappaB via a novel hexanucleotide-binding site. *Nucleic Acids Res* **32**, 3566-80.
- Zelivianski S., Igawa T., Lim S., Taylor R. & Lin M.F. (2002) Identification and characterization of regulatory elements of the human prostatic acid phosphatase promoter. *Oncogene* **21**, 3696-705.
- Zelivianski S., Larson C., Seberger J., Taylor R. & Lin M. (2000) Expression of human prostatic acid phosphatase gene is regulated by upstream negative and positive elements. *Biochim Biophys Acta* **1491**, 123-32.
- Zhang J. (2003) Evolution by gene duplication: an update. *Trends Ecol Evol* **18**, 292-8.
- Zhang J., Rosenberg H.F. & Nei M. (1998) Positive Darwinian selection after gene duplication in primate ribonuclease genes. *PNAS* **95**, 3708-13.
- Zhang X.Q., Lee M.S., Zelivianski S. & Lin M.F. (2001) Characterization of a prostate-specific tyrosine phosphatase by mutagenesis and expression in human prostate cancer cells. *J Biol Chem* **276**, 2544-50.
- Zylka M.J., Sowa N.A., Taylor-Blake B., Twomey M.A., Herrala A., Voikar V. & Vihko P. (2008) Prostatic acid phosphatase is an ectonucleotidase and suppresses pain by generating adenosine. *Neuron* **60**, 111-22.

HUMAN	<b>GGTACC</b> CGCCTTTTGTAGTGTCCCCAGTGAGATCATAGCTGGAGAGGCCAA
CHIMPANZEE	<b>GGTACC</b> CGCCTTTTGTAGTGTCCCCAGTGAGATCATAGCTGGAGAGGCCAA
BONOBO	<b>GGTACC</b> CGCCTTTTGTAGTGTCCCCAGTGAGATCATAGCTGGAGAGGCCAA
GORILLA	<b>GGTACC</b> CGCCTTTTGTAGTGTCCCCAATGAGATCATAGCTGGAGAGGCCAA
ORANGUTAN	<b>GGTACC</b> CGCCTTTTGTAGTGTCCCCAATAAGATCATAGCTGGAGAGGCCAA
HUMAN	GGATGTAACCTCACTCAAAATTTAAGGCTTATGGCCAAAGTTGATGACTAA
CHIMPANZEE	GGATGTAACCTCACTCAAAATTTAAGGCTTATGGCCAAAGTTGATGACTAA
BONOBO	GGATGTAACCTCACTCAAAATTTAAGGCTTATGGCCAAAGTTGATGACTAA
GORILLA	GGATGTAACCTCACTCAAAATTTAAGGCTTATGGCCAAAGTTGATGACTAA
ORANGUTAN	GGATGTAACCTCACTCAAAATTTAAGGCTTATGGCCAAAGGTGATGACTAA
HUMAN	TAATATAGAATTGTAGGTTTGGGCTTTTTGCTTTCCTAAAAAACTCTGTCT
CHIMPANZEE	TAATATAGAATTGTAGGTTTGGGCTTTTTGCTTTCCTAAAAAACTCTGTCT
BONOBO	TAATATAGAATTGTAGGTTTGGGCTTTTTGCTTTCCTAAAAAACTCTGTCT
GORILLA	TAATATAGAATTGTAGGTTTGGGCTTTTTGCTTTCCTAAAAAACTCTGTCT
ORANGUTAN	TAATATAGAATTGTAGGTTTGGGATTTTTGCTTTCCTAAAAAACTCTGTCT
HUMAN	AGGTTTGGAAAGGGACCTCCAATCTAATTTTGAGATACCACAGGTAATGT
CHIMPANZEE	AGGTTTGGAAAGGGACCTCCAATCTAATTTTGAGATACCACAGGTAATGT
BONOBO	AGGTTTGGAAAGGGACCTCCAATCTAATTTTGAGATACCACAGGTAATGT
GORILLA	AGGTTTGGAAAGGGACTTCCAATCTAATTTTGAGATACCACAGGTAATGT
ORANGUTAN	AGGTTTGGAAAGGGACCTCCAATCTAATTTTGAGATACCACAGGTAATGT
HUMAN	CCAATTGTCTTAAGGTGTGACTAAAATTTTTCTTTTGTTTGTTTTTTGT-
CHIMPANZEE	CCAATTGTCTTAAGGTGTGACTAAAATTTTTCTTTTGTTTGTTTTTTGT-
BONOBO	CCAATTGTCTTAAGGTGTGACTAAAATTTTTCTTTTGTTTGTTTTTTGT-
GORILLA	CCAGTTGTCTTACGGTGTGATTAAAACTTTTCTTTTGTTTGTTTTTTGT-
ORANGUTAN	CCAATTGTCTTAAGGTGTGACTAAAATTCCTCTTTTTTGTTTGTTTGTTT
HUMAN	-----TGTTGTTTGTTTGTTTGTTTGCTGTGTTTTGAGATGGAGTCTTG
CHIMPANZEE	-----TGTCGTTTTG-----TTTGTTTGTTTGTTTGTTTTGAGATGGAGTCTTG
BONOBO	-----TGTCGTTTGTTTGTTTGTTTGTTTGTTTGTTTTGAGATGGAGTCTTG
GORILLA	-----TGTTGTTTGTTTGTTTGTTTGTTTGATTGTTTTGAGATGGAGTCTTG
ORANGUTAN	GTTTGTTGTTATTTTTTATTTGTTTGTTTGTTTGTTTTGAGATGGAGTCTTG
HUMAN	CTCTGTCACCTCAGGCTGGAGTGCAGTGGCACAATCTCGGCTCACTGCAAC
CHIMPANZEE	CTCTGTCACCTCAGGCTGGAGTGCAGTGGCACAATCTCGGCTCACTGCAAC
BONOBO	CTCTGTCACCTCAGGCTGGAGTGCAGTGGCACAATCTCGGCTCACTGCAAC
GORILLA	CTCTGTCACCTCAGGCTGGAGTGCAGTGGCACAGTCTCGGCTCACTGCAAC
ORANGUTAN	CTCTGTCACCTCAGGCTGGAGTGCAGTGGCACAATCTCAGCTCACTGCAAC
HUMAN	CTCCACCTCCTGGGTTC AAGTGACTCTCCTGCCTCGCCTCCCAGGTAGCT
CHIMPANZEE	CTCCACCTCCTGGGTTC AAGTGACTCTCCTGCCTCGCCTCCCAGGTAGCT
BONOBO	CTCCACCTCCTGGGTTC AAGTGACTCTCCTGCCTCGCCTCCCAGGTAGCT
GORILLA	CTCCACCTCCTGGGTTC AAGTGACTCTCCTGCCTCGCCTCCCAGGTAGCT
ORANGUTAN	CTCCACCTCCTGGGTTC AAGTGACTCTCCTGCCTCACTTCCCAGGTAGCT
HUMAN	GGGACTATAGGCATGTGCCACCACGCCAGGCTAATTTTTTT-TTTTTTGTA
CHIMPANZEE	GGGACTATAGGCATGTGCCACCACGCCAGGCTAATTTTTTT-TTTTTTGTA
BONOBO	GGGACTATAGGCATGTGCCACCACGCCAGGCTAATTTTTTT--TTTTTTGTA
GORILLA	GGGACTATAGGCAGGTGCCACCACGCCAGGCTAGTTTTTTTTTTTTTTTGT
ORANGUTAN	GGGACTACAGGCATGTGCCACCATGCCACGCTAAAATTT--TTTTTTGT

HUMAN	TTTTTAGTAGAGATGGGTTTTTCACCATGCTGGCCAGGCTGGTCTTAACT
CHIMPANZEE	TTTTTATTAGAGATGGGGTTTTTCACCATGCTGGCCAGGCTGGTCTTAACT
BONOBO	TTTTTATTAGAGATGGGGTTTTTCACCATGCTGGCCAGGCTGGTCTTAACT
GORILLA	TTTTTAGTAGAGATGGGGTTTTTCACCATGCTGGCCAGGCTGGTCTTAACT
ORANGUTAN	TGTTTAGTAGAGATAGGGTTTTTCACCATGCTGGCCAGGCTGGTCTTAACT
HUMAN	CCTGACCTCAAGTGATCTGCCCACCTCGGCCTCCCAAAGTGCTGGGATTA
CHIMPANZEE	CCTGACCTCAAGTGATCTGCCCACCTCAGCCTCCCAAAGTGCTGGGATTA
BONOBO	CCTGACCTCAAGTGATCTGCCCACCTCAGCCTCCCAAAGTGCTGGGATTA
GORILLA	CCTGACCTCAAGTGATCTGCCCACCTCGGCCTCCCAAAGTGCTGGGATTA
ORANGUTAN	CCTGACCTCAAGTGATCTGCCCACCTCGGCCTCCCAAATGCTGGGATTA
HUMAN	CAGACATGAGCCACCACGCCAGCCTAAAATTTTTCAAATAAACTTGCT
CHIMPANZEE	CAGACATGAGCCACCACGCCAGCCTAAAATTTTTCAAATAAACTTGCT
BONOBO	TAGACATGAGCCACCACGCCAGCCTAAAATTTTTCAAATAAACTTGCT
GORILLA	CAGATATGAGCCACCACGCCAGCCTAAAATTTTTCAAATAAACTTGCT
ORANGUTAN	CAGACGTGAGCCACCAGGCCAGCCTAAAATTTTTCAAATAAACTTGCT
HUMAN	GTTACTTTTCTTCTTTAATTTAAAAATACCTCTATAGGCTACATGAGAGT
CHIMPANZEE	GTTACTTTTCTTCTTTAATTTAAAAATACCTCTATAGGCTACATGAGAGT
BONOBO	GTTACTTTTCTTCTTTAATTTAAAAATACCTCTATAGGCTACATGAGAGT
GORILLA	GTTACTTTTCTTCTTTAATTTAAAAATACCTCTATAGGCTACATGAGAGT
ORANGUTAN	CTTACTTCTCTTCTTTAATTTAAAAATACCTCTATAGGCTACATGAGAGT
HUMAN	TTTTGAAAAAGTAGCCATGTTTGGCATGGAAATAGGCCCCAAATATGTCA
CHIMPANZEE	TTTTGAAAAAGTAGCCATGTTTGGCATGGAAATAGGCCCCAAATATGTCA
BONOBO	TTTTGAAAAAGTAGCCATGTTTGGCATGGAAATAGGCCCCAAATATGTCA
GORILLA	TTTTGAAAAAGTAGCCATGTTTGGCATGGAAATAGGTCCCAAATATGTCA
ORANGUTAN	TTTTGAAAAAGTAGCCATGTTTGGCATGGAAATAGGCCCCAAATATGTCA
HUMAN	ACACCTGTATTTTCCTTCCAGATTTTCAGTCCTACAACCTT-GGACTCCCAAC
CHIMPANZEE	ACACCTGTATTTTCCTTCCAGATTTTCAGTCCTACAACCTTTGGACTCCCAAC
BONOBO	ACACCTGTATTTTCCTTCCAGATTTTCAGTCCTACAACCTTTGGACTCCCAAC
GORILLA	ACACCTGTATTTTCCTTCCAAATTTTCAGTCCTACAACCTTTGGACTCCCGAC
ORANGUTAN	ACACCTGTATTTTCCTTCCAGATTTTCAGTCCTACAACCTTTGGACTCCCAAC
HUMAN	AGATTAAGAATCACTGGTCTAAGGT-TTCTAAAGTCAGGGTAGAT-CCCT
CHIMPANZEE	AGATTAAGAATCACTGGTCTAAGGT-TTCTAAAGTCAGGGTAGATCCCTT
BONOBO	AGATTAAGAATCACTGGTCTAAGGT-TTCTAAAGTCAGGGTAGATCCCTT
GORILLA	AGATTAAGAATCACTGGTCTAAGGTTTCTAAAGTCAGGGTAGATCCCTT
ORANGUTAN	AGATTAAGAATCACTGGTCTAAGGT-TTCTAAAGTCAGGGTAGATCCCTT
HUMAN	TTCCAAGTGACAGCCACAAATCCAACGCTGTATCATAACCCTATCTAA--G
CHIMPANZEE	TCCAAGTGACAGCCACAAATCCAACGCTGTATCATAACCCTATCTAAAGA
BONOBO	TCCAAGTGACAGCCACAAATCCAACGCTGTATCATAACCCTATCTAAAGA
GORILLA	TTCCAAGTGACAGCCACAAATCCAACGCTGTATCATAACCCTATCTAAAAA
ORANGUTAN	TCCAAGTGACAGCCACAAATCCAACACTGTATCATA-CCTACCTAAAAA
HUMAN	AAAGATAAAAGTAAACTGGCTTGGCACAGTGGCTCATGCCTGTAATCTCA
CHIMPANZEE	AAAGCTAAAAGTAAACTGGCTTGGCACAGTGGCTCATGCCTGTAATCTCA
BONOBO	AAAGCTAAAAGTAAACTGGCTTGGCACAGTGGCTCATGCCTGTAATCTCA
GORILLA	AAAGATAAAAGTAAACTGGCTTGGCACAGTGGCTCATGCCTGTAATCTCA
ORANGUTAN	AAAGATAAAAGTAAACTGGCAGGGCACGGTGGCTCATGCCTGTAATCTCA

HUMAN	GCACTTTGGGAGGCCAAGGCAGGCGGATCATGCGGTTAGGAGTTGGAGAC
CHIMPANZEE	GCACTTTGGGAGGCCAAGGCAGGCGATCATGCGGTTAGGAGTTGGAGAC
BONOB	GCACTTTGGGAGGCCAAGGCAGGCGGATCATGCGGTTAGGAGTTGGAGAC
GORILLA	GCACTTTGGGAGGCCAAGGCAGGCGGATCATGTGGTTAGGAGTTGGAGAC
ORANGUTAN	GCACTTTGGGAGGCCGAGGCAGGCGATCGTGTGGTTAGGAGTTGGAGAC
HUMAN	CAGCCTAACCAACATGGTGAAACCCCGTCTCTACTAAAAATACAAAAATT
CHIMPANZEE	CAGCCTAACCAACATGGTGAAACCCCGTCTCTACTAAAAATACAAAAATT
BONOB	CAGCCTAACCAACATGGTGAAACCCCGTCTCTACTAAAAATACAAAAATT
GORILLA	CAGCCTAACCAACATGGTGAAACCCCGTCTCTACTAAAAATAAAAAAATT
ORANGUTAN	TAGCCTAACCAACATGGTGAAACCCCGTCTCTACTAAAAATACAAAAATT
HUMAN	CGCCGGGTGTGGTGGTGCACGCCTGTAATCCCAGCTACTCAGGAAGACTG
CHIMPANZEE	CGCCGGGTGTGGTGGTGCACGCCTGTAATCCCAGCTACTCAGG-AGACTG
BONOB	CGCCGGGTGTGGTGGTGCACGCCTGTAATCCCAGCTACTCAGG-AGACTG
GORILLA	CGCCGGGTGTGGTGGTGCACGCCTGTAATCCCAGCTACTCAGG-AGACTG
ORANGUTAN	AGCCGGGTGTGGTGGCGCCCGCTGTAATCCCAGCTACTCAGG-AGGCTG
HUMAN	AAGCAGGAGAATTGCTTGAACCTGGGAGGCGGAGTTTGCAGTGAGCTGAG
CHIMPANZEE	AAGCAGGAGAATTGCTTGAACCTGGGAGGCGGAGTTTGCAGTGAGCCGAG
BONOB	AAGCAGGAGAATTGCTTGAACCTGGGAGGCGGAGTTTGCAGTGAGCCGAG
GORILLA	AAGCAGGAGAATTGCTTGAACCTGGGAGGCGGAGTTTGCAGTGAGCCGAG
ORANGUTAN	AAGCAGGAGAATTGCTTGAACCTGGGAGACGGAGTTTGCAGTGAGCCGAG
HUMAN	ATTGTGCCACTGTGCTCCAGCCTGGGTGACAGAGCAAGACTCTGTCTCAA
CHIMPANZEE	ATTGTGCCACTGCGCTCCAGCCTGGGTGACAGAGCGAGACTCCGTCTCAA
BONOB	ATTGTGCCACTGCGCTCCAGCCTGGGTGACAGAGAGAGACTCCATCTCAA
GORILLA	ATTGTGCCACTGCGCTCCAGCCTGGGTGACAGAGCGAGACTCGGTCTCAA
ORANGUTAN	ATTGTGCCACTGCGCTCCAGCCTGGGTGACAGAGCGAGACTCTGTCTCAA
HUMAN	AAAAAAAAAAAC-----AAAAAACAAAGATAAAAAGTAAACTGAAAACA
CHIMPANZEE	AAAAAAAAAAAAAAAAAAAAAAAAAAAAACAAAGATAAAAAGTAAACTGAAAACA
BONOB	AAAAAAAAAAAAAAAAAAAAAAAAAAAAACAAAGATAA-AAGTAACTGAAAACA
GORILLA	AAAAAA--AAA-----AAAAAACAAAGATAAAAAGTAAACTGAAAACA
ORANGUTAN	AAAAAAAAAA-----AAAAAGATAAAAAGTAAACTGAAAACA
HUMAN	AGGATAGAAATGTCTGCTAACATGCTCTGTGACCTTGATCAACAGCTTTT
CHIMPANZEE	AGAATAGAAATGTCTGCTAACATGCTCTGTGACCTTGATCAACAGCTTTT
BONOB	AGAATAGAAATGTCTGCTAACATGCTCTGTGACCTTGATCAACAGCTTTT
GORILLA	AGAATAGAAATGTCTGCTAACATGCTCTGTGACCTTGATCAACAGCTTTT
ORANGUTAN	AGAATAGAAATGTCTGCTAACATGCTCTGTGACCTTGATCAACAGCTCTC
HUMAN	CCCCTCTTGGCCTTGGTGTCTTAACCTTAAATGGGGAAACTGTGATCTCT
CHIMPANZEE	CCTGCTCTTGGCCTTGGTGTCTTAACCTTAAATGGGGAAACTGTGATCTCT
BONOB	CCTGCTCTTGGCCTTGGTGTCTTAACCTTAAATGGGGAAACTGTGATCTCT
GORILLA	CCCCTCTTGGCCTTGGTGTCTTAACCTTAAATGGGGAAACTGTGATCTCT
ORANGUTAN	CCCCTCTTGGCCTTGGTGTCTTAACCTTAAATGGGGAAACTGTGATCTCT
HUMAN	CTCAGCTCAAAATTTACATGACATTATTTACATACCAAACCTGCCCTACG
CHIMPANZEE	CTCAGCTCAAAATTTACATGACATTATTTACATACCAAACCTGCCCTACG
BONOB	CTCAGCTCAAAATTTACATGACATTATTTACATACCAAACCTGCCCTACG
GORILLA	CTCAGCTCAAAATTTACATGACATTATTTACATACCAAACCTGCCCTACG
ORANGUTAN	CTCAGCTCAAAATTTACATGACATTATTTACATACCAAACCTGCCCTACG
HUMAN	CTTCCTGGAAACTTTACTACTTATGAGTGTGGCTCCTCCTTCCCTTTCAA

CHIMPANZEE	CTTCCTGGAACTTTACTATTTATGAGTGTGGCTCCTCCTTCCCTTTCAA
BONobo	CTTCCTGGAACTTTACTATTTATGAGTGTGGCTCCTCCTTCCCTTTCAA
GORILLA	CTTCCTGGAACTTTACTATTTATGAGTGTGGCTCCTCCTTCCCTTTCAA
ORANGUTAN	CTTCCTGGAACTTTACTATTTATGAGTGTGGCTCCTCCTTCCCTTTCAA
HUMAN	TCCCTTAATTAAATAGCTTCCCCTCTACAGGCTTTTGAAGTGGTAGCAGT
CHIMPANZEE	TCCCTTAATTAAATAGCTTCCCCTCTACAGGCTTTTGAAGTGGTAGCAGT
BONobo	TCCCTTAATTAAATAGCTTCCCCTCTACAGGCTTTTGAAGTGGTAGCAGT
GORILLA	TCCCTTAATTAAATAGCTTCCCCTCTACAGGCTTTTGAAGTGGTAGCAGT
ORANGUTAN	TCCCTTAATTAAATAGCTTCCCCTCTACAGGCTTTTGAAGTGGTAGCAGT
HUMAN	TCCTCCTAACTCCTGCCAGAAACAGCTCTCCTCAACATGAGAGCTGCACC
CHIMPANZEE	TCCTCCTAACTCCTGCCAGAAACAGCTCTCCTCAACATGAGAGCTGCACC
BONobo	TCCTCCTAACTCCTGCCAGAAACAGCTCTCCTCAACATGAGAGCTGCACC
GORILLA	TCCTCCTAACTCCTGCCAGAAACAGCTCTCCTCAACATGAGAGCTGCACC
ORANGUTAN	TCCTCCTAACTCCTGCCAGAAACAGCTCTCCTCAACAAGAGAGCTGCACC
HUMAN	CCTCCTCCTGGCCAGGGCAGCAAGCCTTAGCCTTGGCTTCTTGTTTCTGC
CHIMPANZEE	CCTCCTCCTGGCCAGGGCAGCAAGCCTTAGCCTTGGCTTCTTGTTTCTGC
BONobo	CCTCCTCCTGGCCAGGGCAGCAAGCCTTAGCCTTGGCTTCTTGTTTCTGC
GORILLA	CCTCCTCCTGGCCAGGGCAGCAAGCCTTAGCCTTGGCTTCTTGTTTCTGC
ORANGUTAN	CCTCCTCGTGGCCAGGGCAGCAAGCCTTAGCCTTGGCTTCTTGTTTCTGC
HUMAN	TTTTTTTCTGGCTAGACCGAAGTGTACTAGCCAAGGAGTTGAAGTTTGTG
CHIMPANZEE	TTTTTTTCTGGCTAGACCGAAGTGTACTAGCCAAGGAGTTGAAGTTTGTG
BONobo	TTTTTTTCTGGCTAGACCGAAGTGTACTAGCCAAGGAGTTGAAGTTTGTG
GORILLA	TTTTTTTCTGGCTAGACCGAAGTGTACTAGCCAAGGAGTTGAAGTTTGTG
ORANGUTAN	TTTTTTTCTGGCTAGACCGAAGTGTACTAGCCAAGGAGTTGAAGTTTGTG
HUMAN	ACTTTCCTAAGTAGACTTTTCTCATTGCTTTTTCTTAAACTGTGTTCC
CHIMPANZEE	ACTTTCCTAAGTAGACTTTTCTCATTGCTTTTTCTTAAACTGTGTTCC
BONobo	ACTTTCCTAAGTAGACTTTTCTCATTGCTTTTTCTTAAACTGTGTTCC
GORILLA	ACTTTCCTAAGTAGACTTTTCTCATTGCTTTTTCTTAAACTGTGTTCC
ORANGUTAN	ACTTTCCTAAGTAGACTTTTCTCATTGCTTTTTCTTAAACTGTGTTCC
HUMAN	CAGCAAAGTCTGATAAGGCAAGCGTCAGGTTTCATCTTATCCTTGG----
CHIMPANZEE	CAGCAAAGTCTGATAAGGCAAGCGTCAGCTTTCATCTTATCCTTGGATTG
BONobo	CAGCAAAGTCTGATAAGGCAAGCGTCAGCTTTCATCTTATCCTTGGATTG
GORILLA	CAGCAAAGTCTGATAAGGCAAGCGTCAGGTTTCATCTTATCCTTGG----
ORANGUTAN	CAGCAAAGTCTGATAAGGCAAGCGTCAGGTTTCATCTTACCCTTGG----
HUMAN	-----ATTGTTTCCCAGAGACCAGGCTCTGTTCAAACCTCAGAA
CHIMPANZEE	TTTTATCCTTGGATTGTTTCCCAGAGACCAGGCTCTGTTCAAACCTCAGAA
BONobo	TTTTATCCTTGGATTGTTTCCCAGAGACCAGGCTCTGTTCAAACCTCAGAA
GORILLA	-----ATTGTTTCCCAGAGACCAGGCTCTGTTCAAACCTCAGAA
ORANGUTAN	-----ATTGTTTCCCAGAGACCAGGCTCTGTTCAAACCTCAGAA
HUMAN	AGTGAATAATCAAGTTTTG <b>AAGCTT</b>
CHIMPANZEE	AGTGAATAATCAAGTTTTG <b>AAGCTT</b>
BONobo	AGTGAATAATCAAGTTTTG <b>AAGCTT</b>
GORILLA	AGTGAATAATCAAGTTTTG <b>AAGCTT</b>
ORANGUTAN	AGTGAATAATCAAGTTTTG <b>AAGCTT</b>

**Figure 5-1 ACPP Putative Promoter Alignments**

HUMAN_Seq	CGCCTTTTGTAGTGTCCCCAGTGAGATCATAGCTGGAGAGGCAAGGATGTAAC
Human_UCSC	CGCCTTTTGTAGTGTCCCCAGTGAGATCATAGCTGGAGAGGCAAGGATGTAAC *****
HUMAN_Seq	CACTCAAAATTTAAGGCTTATGGCCAAAGTTGATGACTAATAATATAGAATTGTAGGTTT
Human_UCSC	CACTCAAAATTTAAGGCTTATGGCCAAAGTTGATGACTAATAATATAGAATTGTAGGTTT *****
HUMAN_Seq	GGGCTTTTTGCTTTCTTAAAAAAGTCTGTCAGGTTTGGAAAGGGACCTCCAATCTAATTT
Human_UCSC	GGGCTTTTTGCTTTCTTAAAAAAGTCTGTCAGGTTTGGAAAGGGACCTCCAATCTAATTT *****
HUMAN_Seq	TGAGATACCACAGGTAATGTCCAATTGTCTTAAGGTGTGACTAAAATTTTCTTTTGT
Human_UCSC	TGAGATACCACAGGTAATGTCCAATTGTCTTAAGGTGTGACTAAAATTTTCTTTTGT *****
HUMAN_Seq	GTTTTTGTGTTGTTTGTGTTTGTGTTTGTGCTGTGTTTGTGAGATGGAGTCTTGCTCTGT
Human_UCSC	GTTTTTGTGTTGTTGTTTGTGTTTGTGCTGTGTTTGTGAGATGGAGTCTTGCTCTGT *****
HUMAN_Seq	CACTCAGGCTGGAGTGCAGTGGCACAATCTCGGCTCACTGCAACCTCCACCTCCTGGGTT
Human_UCSC	CACTCAGGCTGGAGTGCAGTGGCACAATCTCGGCTCACTGCAACCTCCACCTCCTGGGTT *****
HUMAN_Seq	CAAGTGACTCTCCTGCCTCGCCTCCCAGGTAGCTGGGACTATAGGCATGTGCCACCACGC
Human_UCSC	CAAGTGACTCTCCTGCCTCGCCTCCCAGGTAGCTGGGACTATAGGCATGTGCCACCACGC *****
HUMAN_Seq	CAGGCTAATTTTTTTTTTTTGTATTTTGTAGTAGAGATGGGTTTCCACATGCTGGCCAGG
Human_UCSC	CAGGCTAATTTTTTTTTTTTGTATTTTGTAGTAGAGATGGGTTTCCACATGCTGGCCAGG *****
HUMAN_Seq	CTGGTCTTAACTCCTGACCTCAAGTGATCTGCCCACCTCGGCCTCCCAAAGTGCTGGGAT
Human_UCSC	CTGGTCTTAACTCCTGACCTCAAGTGATCTGCCCACCTCGGCCTCCCAAAGTGCTGGGAT *****
HUMAN_Seq	TACAGACATGAGCCACCACGCCAGCCTAAAATTTTTCAAATAAACTTGCTGTTACTTT
Human_UCSC	TACAGACATGAGCCACCACGCCAGCCTAAAATTTTTCAAATAAACTTGCTGTTACTTT *****
HUMAN_Seq	TCTTCTTTAATTTAAAAATACCTCTATAGGCTACATGAGAGTTTTTGAAAAAGTAGCCAT
Human_UCSC	TCTTCTTTAATTTAAAAATACCTCTATAGGCTACATGAGAGTTTTTGAAAAAGTAGCCAT *****
HUMAN_Seq	GTTTGGCATGGAAATAGGCCCAAATATGTCAACACCTGTATTTCCTTCCAGATTTTCAAGT
Human_UCSC	GTTTGGCATGGAAATAGGCCCAAATATGTCAACACCTGTATTTCCTTCCAGATTTTCAAGT *****
HUMAN_Seq	CCTACAACCT-TGGACTCCCAACAGATTAAGAATCACTGGTCTAAGGTTTCTAAAGTCAGG
Human_UCSC	CCTACAACCTTTGGACTCCCAACAGATTAAGAATCACTGGTCTAAGGTTTCTAAAGTCAGG ****
HUMAN_Seq	GTAGATCCCTTTC-CAAGTGACAGCCACAAATCCAACGCTGTATCATACCCTATCTAA--
Human_UCSC	GTAGATCCCTTTC-CAAGTGACAGCCACAAATCCAACGCTGTATCATACCCTATCTAAAG *****

HUMAN_Seq	GAAAGATAAAAGTAAACTGGCTTGGCACAGTGGCTCATGCCTGTAATCTCAGCACTTTGG
Human_UCSC	AAAAGATAAAAGTAAACTGGCTTGGCACAGTGGCTCATGCCTGTAATCTCAGCACTTTGG *****
HUMAN_Seq	GAGGCCAAGGCAGGCGGATCATGCGGTTAGGAGTTGGAGACCAGCCTAACCAACATGGTG
Human_UCSC	GAGGCCAAGGCAGGCGGATCATGCGGTTAGGAGTTGGAGACCAGCCTAACCAACATGGTG *****
HUMAN_Seq	AAACCCCGTCTCTACTAAAAATACAAAAATTCGCCGGGTGTGGTGGTGACGCCTGTAAT
Human_UCSC	AAACCCCGTCTCTACTAAAAATACAAAAATTCGCCGGGTGTGGTGGTGACGCCTGTAAT *****
HUMAN_Seq	CCCAGCTACTCAGGAAGACTGAAGCAGGAGAATTGCTTGAACCTGGGAGGCGGAGTTTGC
Human_UCSC	CCCAGCTACTCAGGAAGACTGAAGCAGGAGAATTGCTTGAACCTGGGAGGCGGAGTTTGC *****
HUMAN_Seq	AGTGAGCTGAGATTGTGCCACTGTGCTCCAGCCTGGGTGACAGAGCAAGACTCTGTCTCA
Human_UCSC	AGTGAGCTGAGATTGTGCCACTGTGCTCCAGCCTGGATGACAGAGCAAGACTCTGTCTCA *****
HUMAN_Seq	AAAAAAAAAAAAACAAAAACAAAGATAAAAGTAAACTGAAAACAAGGATAGAAATGTCTG
Human_UCSC	AAAAAAAAAAAAACAAAAACAAAGATAAAAGTAAACTGAAAACAAGGATAGAAATGTCTG *****
HUMAN_Seq	CTAACATGCTCTGTGACCTTGATCAACAGCTTTCCCCGCTCTTGGCCTTGGTGTCTTAAC
Human_UCSC	CTAACATGCTCTGTGACCTTGATCAACAGCTTTCCCCGCTCTTGGCCTTGGTGTCTTAAC *****
HUMAN_Seq	TTAAATGGGGAAACTGTGATCTCTCTCAGCTCAAAATTTACATGACATTATTTACATACC
Human_UCSC	TTAAATGGGGAAACTGTGATCTCTCTCAGCTCAAAATTTACATGACATTATTTACATACC *****
HUMAN_Seq	AAACCTGCCCTACGCTTCCTGGAAACTTTACTACTTATGAGTGTGGCTCCTCCTTCCCTT
Human_UCSC	AAACCTGCCCTACGCTTCCTGGAAACTTTACTACTTATGAGTGTGGCTCCTCCTTCCCTT *****
HUMAN_Seq	TCAATCCCTTAATTAAATAGCTTCCCCTCTACAGGCTTTTGAAGTGGTAGCAGTTCCTCC
Human_UCSC	TCAATCCCTTAATTAAATAGCTTCCCCTCTACAGGCTTTTGAAGTGGTAGCAGTTCCTCC *****
HUMAN_Seq	TAACTCCTGCCAGAAACAGCTCTCCTCAACATGAGAGCTGCACCCCTCCTCCTGGCCAGG
Human_UCSC	TAACTCCTGCCAGAAACAGCTCTCCTCAACATGAGAGCTGCACCCCTCCTCCTGGCCAGG *****
HUMAN_Seq	GCAGCAAGCCTTAGCCTTGGCTTCTTGTTCTGCTTTTTTTCTGGCTAGACCGAAGTGTA
Human_UCSC	GCAGCAAGCCTTAGCCTTGGCTTCTTGTTCTGCTTTTTTTCTGGCTAGACCGAAGTGTA *****
HUMAN_Seq	CTAGCCAAGGAGTTGAAGTTTGTGACTTTGGTAAGTAGACTTTTCTCATTGCTTTTTTCCT
Human_UCSC	CTAGCCAAGGAGTTGAAGTTTGTGACTTTGGTAAGTAGACTTTTCTCATTGCTTTTTTCCT *****
HUMAN_Seq	TAAACTGTGTTCCCAGCAAAGTCTGATAAGGCAAGCGTCAGGTTTCATCTTATCCTTGG
Human_UCSC	TAAACTGTGTTCCCAGCAAAGTCTGATAAGGCAAGCGTCAGGTTTCATCTTATCCTTGG *****
HUMAN_Seq	ATTGTTTCCCAGAGACCAGGCTCTGTTCAAACCTCAGAAAGTGAATAATCAAGTTTTG
Human_UCSC	ATTGTTTCCCAGAGACCAGGCTCTGTTCAAACCTCAGAAAGTGAATAATCAAGTTTTG *****

CHIMPANZEE_Seq Chimpanzee_UCSC	CGCCTTTTGTAGTGTCCCCAGTGAGATCATAGCTGGAGAGGCAAGGATGTAAC CGCCTTTTGTAGTGTCCCCAGTGAGATCATAGCTGGAGAGGCAAGGATGTAAC *****
CHIMPANZEE_Seq Chimpanzee_UCSC	CACTCAAAATTTAAGGCTTATGGCCAAAGTTGATGACTAATAATATAGAATTGTAGGTTT CACTCAAAATTTAAGGCTTATGGCCAAAGTTGATGACTAATAATATAGAATTGTAGGTTT *****
CHIMPANZEE_Seq Chimpanzee_UCSC	GGGCTTTTGTCTTCTTAAAAAAGTCTGTCAGGTTTGAAAGGGACCTCCAATCTAATTT GGGCTTTTGTCTTCTTAAAAAAGTCTGTCAGGTTTGAAAGGGACCTCCAATCTAATTT *****
CHIMPANZEE_Seq Chimpanzee_UCSC	TGAGATACCACAGGTAATGTCCAATTGTCTTAAGGTGTGACTAAAAATTTTCTTTTGT TGAGATACCACAGGTAATGTCCAATTGTCTTAAGGTGTGACTACAATTTTCTTTTGT *****
CHIMPANZEE_Seq Chimpanzee_UCSC	GTTTTTTGTTGTCGTTTG---TTTGTTTTGTGTTGTTTGGAGTGGAGTCTTGCTCTGT GTTTTTTGTTGTCGTTTGTGTTTGTGTTTGTGTTTGGAGTGGAGTCTTGCTCTGT *****
CHIMPANZEE_Seq Chimpanzee_UCSC	CACTCAGGCTGGAGTGCAGTGGCACAATCTCGGCTCACTGCAACCTCCACCTCCTGGGT CACTCAGGCTGGAGTGCAGTGGCACAATCTCGGCTCACTGCAACCTCCACCTCCTGGGT *****
CHIMPANZEE_Seq Chimpanzee_UCSC	CAAGTGACTCTCCTGCCTCGCCTCCCAGGTAGCTGGGACTATAGGCATGTGCCACCACGC CAAGTGACTCTCCTGCCTCGCCTCCCAGGTAGCTGGGACTATAGGCATGTGCCACCACGC *****
CHIMPANZEE_Seq Chimpanzee_UCSC	CAGGCTAATTTTTTTTTTTTGTATTTTTATTAGAGATGGGGTTTCACCATGCTGGCCAGG CAGGCTAATTTTTTTTTTTTGTATTTTTATTAGAGATGGGGTTTCACCATGCTGGCCAGG *****
CHIMPANZEE_Seq Chimpanzee_UCSC	CTGGTCTTAACCTCTGACCTCAAGTGATCTGCCCACCTCAGCCTCCCAAAGTGTGGGAT CTGGTCTTAACCTCTGACCTCAAGTGATCTGCCCACCTCAGCCTCCCAAAGTGTGGGAT *****
CHIMPANZEE_Seq Chimpanzee_UCSC	TACAGACATGAGCCACCACGCCAGCCTAAAATTTTCAAATAAACTTGCTGTACTTT TACAGACATGAGCCACCACGCCAGCCTAAAATTTTCAAATAAACTTGCTGTACTTT *****
CHIMPANZEE_Seq Chimpanzee_UCSC	TCTTCTTTAATTTAAAAATACCTCTATAGGCTACATGAGAGTTTTTGAAAAAGTAGCCAT TCTTCTTTAATTTAAAAATACCTCTATAGGCTACATGAGAGTTTTTGAAAAAGTAGCCAT *****
CHIMPANZEE_Seq Chimpanzee_UCSC	GTTTGGCATGGAATAGGCCCAAATATGTCAACACCTGTATTTCCCTCCAGATTTTCAGT GTTTGGCATGGAATAGGCCCAAATATGTCAACACCTGTATTTCCCTCCAGATTTTCAGT *****
CHIMPANZEE_Seq Chimpanzee_UCSC	CCTACAACCTTTGGACTCCCAACAGATTAAGAATCACTGGTCTAAGGTTTCTAAAGTCAGG CCTACAACCTTTGGACTCCCAACAGATTAAGAATCACTGGTCTAAGGTTTCTAAAGTCAGG *****
CHIMPANZEE_Seq Chimpanzee_UCSC	GTAGATCCCTTTCCCAAGTGACAGCCACAAATCCAACGCTGTATCATACCCTATCTAAAG GTAGATCCCTTTCCCAAGTGACAGCCACAAATCCAACGCTGTATCATACCCTATCTAAAG *****
CHIMPANZEE_Seq Chimpanzee_UCSC	AAAAGCTAAAAGTAACTGGCTTGGCACAGTGGCTCATGCCTGTAATCTCAGCACTTTGG AAAAGCTAAAAGTAACTGGCTTGGCACAGTGGCTCATGCCTGTAATCTCAGCACTTTGG *****
CHIMPANZEE_Seq	GAGGCCAAGGCAGGCAGATCATGCGGTTAGGAGTTGGAGACCAGCCTAACCAACATGGTG



Chimpanzee_UCSC	GAGGCCAAGGCAGGCAGATCATGCGGTTAGGAGTTGGAGACCAGCCTAACCAACATGGTG *****
CHIMPANZEE_Seq Chimpanzee_UCSC	AAACCCCGTCTCTACTAAAAATACAAAAATTCCGCCGGGTGTGGTGGTGCACGCCTGTAAT AAACCCCGTCTCTACTAAAAATACAAAAATTCCGCCGGGTGTGGTGGTGCACGCCTGTAAT *****
CHIMPANZEE_Seq Chimpanzee_UCSC	CCCAGCTACTCAGGAGACTGAAGCAGGAGAATTGCTTGAACCTGGGAGGCGGAGTTTGCA CCCAGCTACTCAGGAGACTGAAGCAGGAGAATTGCTTGAACCTGGGAGGCGGAGTTTGCA *****
CHIMPANZEE_Seq Chimpanzee_UCSC	GTGAGCCGAGATTGTGCCACTGCGCTCCAGCCTGGGTGACAGAGCGAGACTCCGTCTCAA GTGAGCCGAGATTGTGCCACTGCGCTCCAGCCTGGGTGACAGAGCGAGACTCCGTCTCAA *****
CHIMPANZEE_Seq Chimpanzee_UCSC	AAAAAAAAAAAAAAAAAAAAAAAAAAAAAGATAAAAGTAAACTGAAAACAAGAATAGAAA AAAAAAAAAAAAAAAAAAAAAAAAAAAAAGATAAAAGTAAACTGAAAACAAGAATAGAAA *****
CHIMPANZEE_Seq Chimpanzee_UCSC	TGTCTGCTAACATGCTCTGTGACCTTGATCAACAGCTTTCCTGCTCTTGGCCTTGGTGT TGTCTGCTAACATGCTCTGTGACCTTGATCAACAGCTTTCCTGCTCTTGGCCTTGGTGT *****
CHIMPANZEE_Seq Chimpanzee_UCSC	CCTAACTTAAATGGGGAAACTGTGATCTCTCTCAGCTCAAATTTACATGACATTATTTA CCTAACTTAAATGGGGAAACTGTGATCTCTCTCAGCTCAAATTTACATGACATTATTTA *****
CHIMPANZEE_Seq Chimpanzee_UCSC	CATACCAAACCTGCCCTACGCTTCCTGGAACTTTACTATTTATGAGTGTGGCTCCTCCT CATACCAAACCTGCCCTACGCTTCCTGGAACTTTACTATTTATGAGTGTGGCTCCTCCT *****
CHIMPANZEE_Seq Chimpanzee_UCSC	TCCCTTTCAATCCCTTAATTAATAGCTTCCCCTCTACAGGCTTTTGAAGTGGTAGCAGT TCCCTTTCAATCCCTTAATTAATAGCTTCCCCTCTACAGGCTTTTGAAGTGGTAGCAGT *****
CHIMPANZEE_Seq Chimpanzee_UCSC	TCCTCCTAACTCCTGCCAGAAACAGCTCTCCTCAACATGAGAGCTGCACCCCTCCTCCTG TCCTCCTAACTCCTGCCAGAAACAGCTCTCCTCAACATGAGAGCTGCACCCCTCCTCCTG *****
CHIMPANZEE_Seq Chimpanzee_UCSC	GCCAGGGCAGCAAGCCTTAGCCTTGCTTCTTGTTCTGCTTTTTTTCTGGCTAGACCGA GCCAGGGCAGCAAGCCTTAGCCTTGCTTCTTGTTCTGCTTTTTTTCTGGCTAGACCGA *****
CHIMPANZEE_Seq Chimpanzee_UCSC	AGTGTACTAGCCAAGGAGTTGAAGTTTGTGACTTTGGTAAGTAGACTTTTCTCATTGCTT AGTGTACTAGCCAAGGAGTTGAAGTTTGTGACTTTGGTAAGTAGACTTTTCTCATTGCTT *****
CHIMPANZEE_Seq Chimpanzee_UCSC	TTTCCTTAAACTGTGTTCCCAGCAAAGTCTGATAAGGCAAGCGTCAGCTTTCATCTTAT TTTCCTTAAACTGTGTTCCCAGCAAAGTCTGATAAGGCAAGCGTCAGCTTTCATCTTAT *****
CHIMPANZEE_Seq Chimpanzee_UCSC	CCTTGGATTGTTTTATCCTTGATTGTTTCCCAGAGACCAGGCTCTGTTCAAACCTCAGAA CCTTGGATTGTTTTATCCTTGATTGTTTCCCAGAGACCAGGCTCTGTTCAAACCTCAGAA *****
CHIMPANZEE_Seq Chimpanzee_UCSC	AGTGAATAATCAAGTTTTG AGTGAATAATCAAGTTTTG *****

---

Bonobo_Seq	CGCCTTTTGTAGTGTCCCCAGTGAGATCATAGCTGGAGAGGCAAGGATGTAAC
------------	---

Chimpanzee_UCSC	CGCCTTTTGTAGTGTCCCCAGTGAGATCATAGCTGGAGAGGCAAGGATGTA *****
Bonobo_Seq Chimpanzee_UCSC	CACTCAAAATTTAAGGCTTATGGCCAAAGTTGATGACTAATAATATAGAATTGTAGGTTT CACTCAAAATTTAAGGCTTATGGCCAAAGTTGATGACTAATAATATAGAATTGTAGGTTT *****
Bonobo_Seq Chimpanzee_UCSC	GGGCTTTTGTCTTCTTAAAAAATCTGTCTAGGTTTGGAAGGGACCTCCAATCTAATTT GGGCTTTTGTCTTCTTAAAAAATCTGTCTAGGTTTGGAAGGGACCTCCAATCTAATTT *****
Bonobo_Seq Chimpanzee_UCSC	TGAGATACCACAGGTAATGTCCAATTGTCTTAAGGTGTGACTAAAAATTTTCTTTTGTTT TGAGATACCACAGGTAATGTCCAATTGTCTTAAGGTGTGACTACAATTTTCTTTTGTTT *****
Bonobo_Seq Chimpanzee_UCSC	GTTTTTGTGTGCTGTTGTTTGTTTGTTTGTTTGTTTGTTTGAGATGGAGTCTTGCTCTGT GTTTTTGTGTGCTGTTGTTTGTTTGTTTGTTTGTTTGTTTGAGATGGAGTCTTGCTCTGT *****
Bonobo_Seq Chimpanzee_UCSC	CACTCAGGCTGGAGTGCAGTGGCACAATCTCGGCTCACTGCAACCTCCACCTCCTGGGT CACTCAGGCTGGAGTGCAGTGGCACAATCTCGGCTCACTGCAACCTCCACCTCCTGGGT *****
Bonobo_Seq Chimpanzee_UCSC	CAAGTGACTCTCCTGCCTCGCCTCCCAGGTAGCTGGGACTATAGGCATGTGCCACCACGC CAAGTGACTCTCCTGCCTCGCCTCCCAGGTAGCTGGGACTATAGGCATGTGCCACCACGC *****
Bonobo_Seq Chimpanzee_UCSC	CAGGCTAATTTTTT-TTTTTGTATTTTTATTAGAGATGGGGTTTCACCATGCTGGCCAGG CAGGCTAATTTTTTTTTTTTTGTATTTTTATTAGAGATGGGGTTTCACCATGCTGGCCAGG *****
Bonobo_Seq Chimpanzee_UCSC	CTGGTCTTAACTCCTGACCTCAAGTGATCTGCCCATCTCAGCCTCCCAAAGTGCTGGGAT CTGGTCTTAACTCCTGACCTCAAGTGATCTGCCCATCTCAGCCTCCCAAAGTGCTGGGAT *****
Bonobo_Seq Chimpanzee_UCSC	TATAGACATGAGCCACCACGCCAGCCTAAAATTTTTCAAATAAACTTGCTGTTACTTT TACAGACATGAGCCACCACGCCAGCCTAAAATTTTTCAAATAAACTTGCTGTTACTTT ** *****
Bonobo_Seq Chimpanzee_UCSC	TCTTCTTTAATTTAAAAATACCTCTATAGGCTACATGAGAGTTTTTGAAAAAGTAGCCAT TCTTCTTTAATTTAAAAATACCTCTATAGGCTACATGAGAGTTTTTGAAAAAGTAGCCAT *****
Bonobo_Seq Chimpanzee_UCSC	GTTTGGCATGGAAATAGGCCCCAAATATGTCAACACCTGTATTTCTTCCAGATTTTCAGT GTTTGGCATGGAAATAGGCCCCAAATATGTCAACACCTGTATTTCTTCCAGATTTTCAGT *****
Bonobo_Seq Chimpanzee_UCSC	CCTACAACCTTTGGACTCCCAACAGATTAAGAATCACTGGTCTAAGGTTTCTAAAGTCAGG CCTACAACCTTTGGACTCCCAACAGATTAAGAATCACTGGTCTAAGGTTTCTAAAGTCAGG *****
Bonobo_Seq Chimpanzee_UCSC	GTAGATCCCTTTCCCAAGTGACAGCCACAAATCCAACGCTGTATCATACCCTATCTAAAG GTAGATCCCTTTCCCAAGTGACAGCCACAAATCCAACGCTGTATCATACCCTATCTAAAG *****
Bonobo_Seq Chimpanzee_UCSC	AAAAGCTAAAAGTAACTGGCTTGGCACAGTGGCTCATGCCTGTAATCTCAGCACTTTGG AAAAGCTAAAAGTAACTGGCTTGGCACAGTGGCTCATGCCTGTAATCTCAGCACTTTGG *****
Bonobo_Seq Chimpanzee_UCSC	GAGGCCAAGGCAGGCAGATCATGCGGTTAGGAGTTGGAGACCAGCCTAACCAACATGGTG GAGGCCAAGGCAGGCAGATCATGCGGTTAGGAGTTGGAGACCAGCCTAACCAACATGGTG

```

*****

Bonobo_Seq      AAACCCCGTCTCTACTAAAAATACAAAAATTGCGCCGGGTGTGGTGGTGCACGCCTGTAAT
Chimpanzee_UCSC AAACCCCGTCTCTACTAAAAATACAAAAATTGCGCCGGGTGTGGTGGTGCACGCCTGTAAT
*****

Bonobo_Seq      CCCAGCTACTCAGGAGACTGAAGCAGGAGAATTGCTTGAACCTGGGAGGCGGAGTTTGCA
Chimpanzee_UCSC CCCAGCTACTCAGGAGACTGAAGCAGGAGAATTGCTTGAACCTGGGAGGCGGAGTTTGCA
*****

Bonobo_Seq      GTGAGCCGAGATTGTGCCACTGCGCTCCAGCCTGGGTGACAGAGAGAGACTCCATCTCAA
Chimpanzee_UCSC GTGAGCCGAGATTGTGCCACTGCGCTCCAGCCTGGGTGACAGAGCGAGACTCCGTCTCAA
*****

Bonobo_Seq      AAAAAAAAAAAAAAAAAAAAAAAAAACAAAGATAAAAGTAA-CTGAAACAAGAATAGAAA
Chimpanzee_UCSC AAAAAAAAAAAAAAAAAAAAAAAAAACAAAGATAAAAGTAAACTGAAACAAGAATAGAAA
*****

Bonobo_Seq      TGTCTGCTAACATGCTCTGTGACCTTGATCAACAGCTTTCCTGCTCTTGGCCTTGGTGT
Chimpanzee_UCSC TGTCTGCTAACATGCTCTGTGACCTTGATCAACAGCTTTCCTGCTCTTGGCCTTGGTGT
*****

Bonobo_Seq      CCTAACTTAAATGGGGAAACTGTGATCTCTCTCAGCTCAAATTTACATGACATTATTTA
Chimpanzee_UCSC CCTAACTTAAATGGGGAAACTGTGATCTCTCTCAGCTCAAATTTACATGACATTATTTA
*****

Bonobo_Seq      CATACCAAACCTGCCCTACGCTTCCTGGAACTTTACTATTTATGAGTGTGGCTCCTCCT
Chimpanzee_UCSC CATACCAAACCTGCCCTACGCTTCCTGGAACTTTACTATTTATGAGTGTGGCTCCTCCT
*****

Bonobo_Seq      TCCCTTTCAATCCCTTAATTAATAGCTTCCCCTCTACAGGCTTTTGAAGTGGTAGCAGT
Chimpanzee_UCSC TCCCTTTCAATCCCTTAATTAATAGCTTCCCCTCTACAGGCTTTTGAAGTGGTAGCAGT
*****

Bonobo_Seq      TCCTCCTAACTCCTGCCAGAAACAGCTCTCCTCAACATGAGAGCTGCACCCCTCCTCCTG
Chimpanzee_UCSC TCCTCCTAACTCCTGCCAGAAACAGCTCTCCTCAACATGAGAGCTGCACCCCTCCTCCTG
*****

Bonobo_Seq      GCCAGGGCAGCAAGCCTTAGCCTTGGCTTCTTGTCTGCTTTTTTCTGGCTAGACCGA
Chimpanzee_UCSC GCCAGGGCAGCAAGCCTTAGCCTTGGCTTCTTGTCTGCTTTTTTCTGGCTAGACCGA
*****

Bonobo_Seq      AGTGTAAGTACTAGCAAGGAGTTGAAGTTTGTGACTTTGGTAAGTAGACTTTTCTCATTTGCTT
Chimpanzee_UCSC AGTGTAAGTACTAGCAAGGAGTTGAAGTTTGTGACTTTGGTAAGTAGACTTTTCTCATTTGCTT
*****

Bonobo_Seq      TTTCCCTTAAACTGTGTTCCCAGCAAAGTCTGATAAGGCAAGCGTCAGCTTTTCATCTTAT
Chimpanzee_UCSC TTTCCCTTAAACTGTGTTCCCAGCAAAGTCTGATAAGGCAAGCGTCAGCTTTTCATCTTAT
*****

Bonobo_Seq      CCTTGGATTGTTTTATCCTTGGATTGTTTCCCAGAGACCAGGCTCTGTTCAAACCTCAGAA
Chimpanzee_UCSC CCTTGGATTGTTTTATCCTTGGATTGTTTCCCAGAGACCAGGCTCTGTTCAAACCTCAGAA
*****

Bonobo_Seq      AGTGAATAATCAAGTTTTG
Chimpanzee_UCSC AGTGAATAATCAAGTTTTG
*****

```

---

```

Gorilla_Seq      TTTTGTAGTGTCCCCAATGAGATCATAGCTGGAGAGGCAAGGATGTAAC
Gorilla_UCSC      TTTTGTAGTGTCCCCAATGAGATCATAGCTGGAGAGGCAAGGATGTAAC

```

[illegible]

Gorilla\_Seq TGAAACCCCGTCTCTACTAAAAATAAAAAATTCGCCGGGTGTGGTGGTGCACGCCTGTA  
Gorilla\_UCSC NNN

Gorilla\_Seq ATCCCAGCTACTCAGGAGACTGAAGCAGGAGAATTGCTTGAACCTGGGAGGCGGAGTTTG  
Gorilla\_UCSC NNN

Gorilla\_Seq CAGTGAGCCGAGATTGTGCCACTGCGCTCCAGCCTGGGTGACAGAGCGAGACTCGGTCTC  
Gorilla\_UCSC NNN

Gorilla\_Seq AAAAAAAAAAAAAAAAAACAAAGATAAAAGTAAACTGAAAACAAGAATAGAAATGTCTGC  
Gorilla\_UCSC NAAAAAAAAAAAAAAAAACAAAGATAAAAGTAAACTGAAAACAAGAATAGAAATGTCTGC  
\*\*\*\*\*

Gorilla\_Seq TAACATGCTCTGTGACCTTGATCAACAGCTTTCCCCGCTCTTGGCCTTGGTGTCTTA  
Gorilla\_UCSC TAACATGCTCTGTGACCTTGATCAACAGCTTTCCCCGCTCTTGGCCTTGGTGTCTTA  
\*\*\*\*\*

Gorilla\_Seq TAAATGGGGAAACTGTGATCTCTCTCAGCTCAAAATTTACATGACATTATTTACATACCA  
Gorilla\_UCSC TAAATGGGGAAACTGTGATCTCTCTCAGCTCAAAATTTACATGACATTATTTACATACCA  
\*\*\*\*\*

Gorilla\_Seq AACCTGCCCTACGCTTCCTGGAACTTTACTATTTATGAGTGTGGCTCCTCCTTCCCTTT  
Gorilla\_UCSC AACCTGCCCTACGCTTCCTGGAACTTTACTATTTATGAGTGTGGCTCCTCCTTCCCTTT  
\*\*\*\*\*

Gorilla\_Seq CAATCCCTTAATTAAATAGCTTCCCCTCTACAGGCTTTTGAAGTGGTAGCAGTTCCTCCT  
Gorilla\_UCSC CAATCCCTTAATTAAATAGCTTCCCCTCTACAGGCTTTTGAAGTGGTAGCAGTTCCTCCT  
\*\*\*\*\*

Gorilla\_Seq AACTCCTGCCAGAAACAGCTCTCCTCAACATGAGAGCTGCACCCCTCCTCCTGGCCAGGG  
Gorilla\_UCSC AACTCCTGCCAGAAACAGCTCTCCTCAACATGAGAGCTGCACCCCTCCTCCTGGCCAGGG  
\*\*\*\*\*

Gorilla\_Seq CAGCAAGCCTTAGCCTTGGCTTCTTGTCTGCTTTTTTTCTGGCTAGACCGAAGTGTAC  
Gorilla\_UCSC CAGCAAGCCTTAGCCTTGGCTTCTTGTCTGCTTTTTTTCTGGCTAGACCGAAGTGTAC  
\*\*\*\*\*

Gorilla\_Seq TAGCCAAGGAGTTGAAGTTTGTGACTTTGGTAAGTAGACTTTTCTCATTGCTTTTTCTT  
Gorilla\_UCSC TAGCCAAGGAGTTGAAGTTTGTGACTTTGGTAAGTAGACTTTTCTCATTGCTTTTTCTT  
\*\*\*\*\*

Gorilla\_Seq AAAACTGTGTTCCCAGCAAAGTCTGATAAGGCAAGCGTCAGGTTTCATCTTATCCTTGGA  
Gorilla\_UCSC AAAACTGTGTTCCCAGCAAAGTCTGATAAGGCAAGCGTCAGGTTTCATCTTATCCTTGGA  
\*\*\*\*\*

Gorilla\_Seq TTGTTTCCCAGAGACCAGGCTCTGTTCAAACCTCAGAAAGTGAATAATCAAGTTTTG  
Gorilla\_UCSC TTGTTTCCCAGAGACCAGGCTCTGTTCAAACCTCAGAAAGTGAATAATCAAGTTTTG  
\*\*\*\*\*

---

ORANGUTAN\_Seq CGCCTTTTGTAGTGTCCCAATAAGATCATAGCTGGAGAGGCAAGGATGTAAC  
Orangutan\_UCSC CGCCTTTTGTAGTGTCCCAATAAGATCATAGCTGGAGAGGCAAGGATGTAAC  
\*\*\*\*\*

ORANGUTAN\_Seq CACTCAAAATTTAAGGCTTATGGCCAAAGGTGATGACTAATAATATAGAATTGTAGGTTT  
Orangutan\_UCSC CACTCAAAATTTAAGGCTTATGGCCAAAGGTGATGACTAATAATATAGAATTGTAGGTTT  
\*\*\*\*\*

ORANGUTAN_Seq Orangutan_UCSC	GGGATTTTGTCTTTCTTAAAAAAGTCTGTCAGGTTTGGAAAGGGACCTCCAATCTAATTT GGGATTTTGTCTTTCTTAAAAAAGTCTGTCAGGTTTGGAAAGGGACCTCCAATCTAATTT *****
ORANGUTAN_Seq Orangutan_UCSC	TGAGATACCACAGGTAATGTCCAATTGTCTTAAGGTGTGACTAAAATTCTTCTTTTGT TGAGATACCACAGGTAATGTCCAATTGTCTTAAGGTGTGACTAAAATTCTTCTTTTGT ***** *
ORANGUTAN_Seq Orangutan_UCSC	TTGTTTGTGTTTGTGTTTATTTTATTTGTTTGTGTTTGTGTTTGTGATGGAGTCTTG G-----TTTGTGTTGTTTGTGTTTGTGTTTGTGTTTGTGATGGAGTCTTG *****
ORANGUTAN_Seq Orangutan_UCSC	CTCTGTCACTCAGGCTGGAGTGCAGTGGCACAATCTCAGCTCACTGCAACCTCCACCTCC CTCTGTCACTCAGGCTGGAGTGCAGTGGCACAATCTCAGCTCACTGCAACCTCCACCTCC *****
ORANGUTAN_Seq Orangutan_UCSC	TGGGTTCAAGTGACTCTCCTGCCTCACTTCCCAGGTAGCTGGGACTACAGGCATGTGCCA TGGGTTCAAGTGACTCTCCTGCCTCACTTCCCAGGTAGCTGGGACTACAGGCATGTGCCA *****
ORANGUTAN_Seq Orangutan_UCSC	CCATGCCACGCTAAAATTTTTTTGTATGTTTAGTAGAGATAGGGTTTCACCATGCTGGC CCATGCCACGCTAAAATTTTTTTGTATGTTTAGTAGAGATAGGGTTTCACCATGCTGGC *****
ORANGUTAN_Seq Orangutan_UCSC	CAGGCTGGTCTTAACTCCTGACCTCAAGTGATCTGCCACCTCGGCTCCCAAAATGCTG CAGGCTGGTCTTGAACCTGACCAAGTGATCTGCCACCTCGGCTCCCAAAATGCTG ***** *
ORANGUTAN_Seq Orangutan_UCSC	GGATTACAGACGTGAGCCACCAGGCCAGCCTAAAATTTTTTCAAATAAACTTGCTCTTA GGATTACAGACGTGAGCCACCAGGCCAGCCTAAAATTTTTTCAAATAAACTTGCTCTTA *****
ORANGUTAN_Seq Orangutan_UCSC	CTTCTCTTCTTTAATTTAAAAATACCTCTATAGGCTACATGAGAGTTTTTGAAAAAGTAG CTTCTCTTCTTTAATTTAAAAATACCTCTATAGGCTACATGAGAGTTTTTGAAAAAGTAG *****
ORANGUTAN_Seq Orangutan_UCSC	CCATGTTTGGCATGGAAATAGGCCCAAATATGTCAACACCTGTATTTCTTCCAGATTT CCATGTTTGGCATGGAAATAGGCCCAAATATGTCAACACCTGTATTTCTTCCAGATTT *****
ORANGUTAN_Seq Orangutan_UCSC	CAGTCCTACAACCTTGGACTCCCAACAGATTAAGAATCACTGGTCTAAGGTTTCTAAAGT CAGTCCTACAACCTTGGACTCCCAACAGATTAAGAATCACTGGTCTAAGGTTTCTAAAGT *****
ORANGUTAN_Seq Orangutan_UCSC	CAGGGTAGATCCCTTTCCCAAGTGACAGCCACAAATCCAACACTGTATCATACCTACCTA CAGGGTAGATCCCTTTCCCAAGTGACAGCCACAAATCCAACACTGTATCATACCTACCTA *****
ORANGUTAN_Seq Orangutan_UCSC	AAAAAAGATAAAAGTAACTGGCAGGGCACGGTGGCTCATGCCTGTAATCTCAGCACTT AAAAAAGATAAAAGTAACTGGCAGGGCACGGTGGCTCATGCCTGTAATCTCAGCACTT *****
ORANGUTAN_Seq Orangutan_UCSC	TGGGAGGCCGAGGCAGGCAGATCGTGTGGTTAGGAGTTGGAGACTAGCCTAACCAACATG TGGGAGGCCGAGGCAGGCAGATCGTGTGGTTAGGAGTTGGAGACTAGCCTAACCAACATG *****
ORANGUTAN_Seq Orangutan_UCSC	GTGAAACCCCGTCTCTACTAAAAATACAAAAATTAGCCGGGTGTGGTGGCGCCCGCTGT GTGAAACCCCGTCTCTACTAAAAATACAAAAATTAGCCGGGTGTGGTGGCGCCCGCTGT *****
ORANGUTAN_Seq	AATCCCAGCTACTCAGGAGGCTGAAGCAGGAGAATTGCTTGAACCTGGGAGACGGAGTTT

Orangutan_UCSC	AATCCCAGCTACTCAGGAGGCTGAAGCAGGAGAATTGCTTGAACCTGGGAGACGGAGTTT *****
ORANGUTAN_Seq Orangutan_UCSC	GCAGTGAGCCGAGATTGTGCCACTGCGCTCCAGCCTGGGTGACAGAGCGAGACTCTG--- GCAGTGAGCCGAGATTGTGCCACTGCGCTCCAGCCTGGGTGACAGAGCGAGACTCNNNNN *****
ORANGUTAN_Seq Orangutan_UCSC	----- NN
ORANGUTAN_Seq Orangutan_UCSC	----- NN
ORANGUTAN_Seq Orangutan_UCSC	----- NN
ORANGUTAN_Seq Orangutan_UCSC	----- NN
ORANGUTAN_Seq Orangutan_UCSC	-----TCTCAAAAAAAAAAAAAA NN *****
ORANGUTAN_Seq Orangutan_UCSC	AAAAAGATAAAAGTAACTGAAAACAAGAATAGAAATGTCTGCTAACATGCTCTGTGACC AAAAAGATAAAAGTAACTGAAAACAAGAATAGAAATGTCTGCTAACATGCTCTGTGACC *****
ORANGUTAN_Seq Orangutan_UCSC	TTGATCAACAGCTCTCCCGCTCTTGGCCTTGGTGTCTAACTTAAATGGGGAAACTGTG TTGATCAACAGCTTTCCCGCTCTTGGCCTTGGTGTCTAACTTAAATGGGGAAACTGTG *****
ORANGUTAN_Seq Orangutan_UCSC	ATCTCTCTCAGCTCAAAATTTACATGACATTATTTACATACCAAACCTGCCCTACGCTTC ATCTCTCTCAGCTCAAAATTTACATGACATTATTTACATACCAAACCTGCCCTACGCTTC *****
ORANGUTAN_Seq Orangutan_UCSC	CTGGAACTTTACTATTTATGAGTGTGGCTCCTCCTTCCCTTTCAATCCCTTAATTAAAT CTGGAACTTTACTATTTATGAGTGTGGCTCCTCCTTCCCTTTCAATCCCTTAATTAAAT *****
ORANGUTAN_Seq Orangutan_UCSC	AGCTTCCCCTCTACAGGCTTTTGAAGTGGTAGCAGTTCCTCCTAACTCCTGCCAGAAACA AGCTTCCCCTCTACAGGCTTTTGAAGTGGTAGCAGTTCCTCCTAACTCCTGCCAGAAACA *****
ORANGUTAN_Seq Orangutan_UCSC	GCTCTCTCAACAAGAGAGCTGCACCCCTCCTCGTGGCCAGGGCAGCAAGCCTTAGCCTT GCTCTCTCAGCATGAGAGCTGCACCCCTCCTCGTGGCCAGGGCAGCAAGCCTTAGCCTT *****
ORANGUTAN_Seq Orangutan_UCSC	GGCTTCTTGTCTTCTGCTTTTTTTCTGGCTAGACCGAAGTGACTAGCCAAGGAGTTGAAG GGCTTCTTGTCTTCTGCTTTTTTTCTGGCTAGACCGAAGTGACTAGCCAAGGAGTTGAAG *****
ORANGUTAN_Seq Orangutan_UCSC	TTTGTGACTTTGGTAAAGTAGACTTTTCTCATTGCTTTTTCTTAAACTGTGTTCCAGC TTTGTGACTTTGGTAAAGTAGACTTTTCTCATTGCTTTTTCTTAAACTGTGTTCCAGC *****

```

ORANGUTAN_Seq      AAAGTCTGATAAGGCAAGCGTCAGGTTTCATCTTACCCTTGGATTGTTTCCCAGAGACCA
Orangutan_UCSC     AAAGTCTGATAAGGCAAGCGTCAGGTTTCATCTTACCCTTGGATTGTTTCCCAGAGACCA
*****

ORANGUTAN_Seq      GGCTCTGTTCAAAC TCAGAAAGTGAATAATCAAGTTTG
Orangutan_UCSC     GGCTCTGTTCAAAC TCAGAAAGTGAATAATCAAGTTTG
*****

```

**Figure 5-2 *ACPP* Regulatory Region Genomic Sequence Aligned to UCSC Sequence**

The genomic sequence of the *ACPP* regulatory region used to create the experimental constructs aligned to the UCSC sequence for each species.



TAGTCATGTGGTGGATTTCGGTCAGGAGTTCGAGACAGCCTGACCACATGGTGAAACTCTGTCTCTACTAAA  
AAAAAAAATAGAAAAATTAGCCAGGCGTGGTGGCATGCAGCACCTGTAATCCCAGCTACTCAGGAGGCTG  
AGGCAGGAGAATCACTTGAACCCAGGAGGCAGAGGTTGCAGTGAGCTGAGATTGCACCACTGCACTCCAGC  
CTGGGTGACAGAGTGAGACTCCGTCTCAAAAAAAAAAATTTTTTTTTTTTTTTTTTTTTTTTTTTGTAGAGA  
TGGATCTTGCTTTGTTTCTCTGGTTGGCCTTGAACCTCGGCTTCAAGTGATCCTCCTACCTTGGCCTCGG  
AAAGTGTTGGGATTACAGGCGTGAGCCACCATGACTGACCTGTCGTTTAATCTTGAGATACATAAACCTG  
GCTCCTAAAGGCTAAATATTTTGTGGAGAAGGGGCATTGGATTTTGCATGAGGATGATTCTGACCTGGG  
AGGGCAGGTCAGCAGGCATCTCTGTTGCACAGATAGAGTGACAGGTCTGGAGAACAAGGAGTGGGGGGT  
TATTGGAATTCCACATTGTTTGTCTGCACGTTGGATTTTGAAATGCTAGGGAACCTTGGGAGACTCATATT  
TCTGGGCTAGAGTATCTGTGGACCATAAGATCTTTTTATGATGACAGTAGCAATGTATCTGTGGAGCTGG  
ATTCTGGGTTGGGAGTGCAAGGAAAAGAGAGTACTAAATGCCAAGACATCTATTTTCAGGAGCATGAGAAA  
TAAAAGTTCTAGTTTCTGGGCTCAGAGTGGTGCAGGGATCAGGGAGTCTCACAATCCCCTGAGTGCTGGTG  
TTTTAGGGCACACTGGGTCTTGGAGTGCAAAGGATCTAGGCACGTGAGGCTTTGTATGAAGAATCGGGGA  
TCGTACCCACCCCCTGTTTCTGTTTCATCCTGGGCCTGTCTCCTCTGCCTTTGTCCCCTAGATGAAGTCTCCA  
TGAGGTACAGGGCCTGGTGCATCCAGGGTGATCTAGTAATTACAGAACAGCAAGTGCTAGCTCTCCCTTCC  
CTTCCACAGCTCTGGGTGTGGGAGGGGGTGTCCAGCCTCCAGCAGCATGGGGAGGGCCTTGGTCAGCCTC  
TGGGTGCCAGCAGGGCAGGGGCGGAGTCCTGGGGAGTGAAGGTTTTATAAGGCTCCTGGGGGAGGCTCCCC  
AGCCCCAAGCTTACCACCTGCACCCGGAGAGCTGTGTACCATGTGGGTCTGGTTGTCTTCTCACCTGT  
CTGTGACGTGGATTGGTGAGAGGGGCCATGGCTGGGGGGATGCAGGAGAGGGAGCCAGTCTGACTGTCA  
AGCTGAGGCTCTTTCCCCCCCACCCAGCACCCcAGCCCAGACAGGGAGCTGGGCTCTTTTCTGTCTCTTCCC  
AGCCCCACTCCAAGCCCGTACCCCCAGCCCCTCCATATTGCAACAGTCTCCTCACTCCCACACCAGGTCCCCGC  
TCCCTCCCACCTACCCCAGAACTTTCTCCCCATTGCCAGCCAACTCCCTGCTCCCAGCTGCTTTACTAAAG  
GGGAAGTTCTGGGCATCTCCGTGTTTCTCTTTGTGGGGCTCAAAACCTCCAAGGACCTCTCTCAATGCCA  
TTGGTTCCTTGGACCCTATCACTGGTCCACCTCCTGAGCCCCTCAATCCTATCACAGTCTACTGAGTTTTCC  
CATTGAGCTGTGAGTGTCCAACCCTATCCCAGAGACCTTGATGCTTGGCCTCCCAATCTTGCCCTAGGATA  
CCCAGATGCCAACCAGACACCTCCTTCTTCTAGCCTGGCTATCTGGCCTGAGACAACAAATGGGTCCCTC  
AGTCTGGCAATGGGACTCTGAGAACTCCTCATTCCCTGACTCTTAGCCACAGATTCTTCATTGAGTGGCGC  
ACATTTTCCTTAGGAAAACTTGAGCATCCCCAGCCTCATTCCCTGACTCTTAGCCCCAGACTCTTCATTC  
AGTGGCGCACATTTTCCTTAGGAAAAACATGAGCATCCCCAGCCACAACCTGCCAGAGCTCTGATTCCCTAA  
ATCTGCATCCTTTTCAACCTAAAAACAAAAAGAAAAACAAATAAAACAAAACCAACTCAGACCAGAACT  
GTTTTCTCAACCTGGGACTTCCTAAACTTTCCAAAACCTTCCTCTTCCAGCAACTGAAGCTCGCCATAAGG  
CACTTATCCCTGGTTCCTAGTACCCCTTATCCCTCAGAATCCACAACCTTGTACCAAGTTTCCCTTCTCCCA

GTCCAAGACCCCAAATCACCACAAAGGACCCGATCCCCAGACTCAAGATATGGTCTGGGCACTGTCTTGTG  
TCTCCTACCCTGATCCCTGGGGTCAACTCTGCTCCCAGAGCATGAAGCCTCTCCACCAGCACCAGCCACCAA  
CCTGCAAACCTAGGGAAGATTGACAGAATTCCCAGCCTTTCCCAGCTCCCTCTGCCCATGTCCCAGGACTC  
CCAGCCTTGGTTCTCTGCCCCCATGTCTTTTCAAACCCACATCCTAAATCCGTCTCCTATCTGAGTCCCCCA  
GTTCTCTCTGTCAACCCTGATCCCCCTGATCTAGCACCCCTCTGCAGGCGCTGCACCCCTCATCCTGTCTC  
GGATTGTGGGAGGCTGGGAGTGCGAGAAGCATTCCCAACCCTGGCAGGTGCTTGTGGCCTCTCACGGCAGG  
GCAGTCTGCGGTGGTGTCTGCTGTCACCCCACTGAGTGGGTCCTCACAGCTGCCACTGCATCAGGAAGTGAGT  
AGGGGCCTGGGGTCTGGGGAGGAGGTGTCTGTGTCCCAGAGGAATAACAGCTGGGCATTTTCCCCAGGAT  
AACCTCTAAGGCGAGCCTTGGGACTGGGGGAGAGAGGGAAGGTTCTGGTTCAGGTCACATGGGGAGGCAG  
GGTTGGGGCTGGACCACCCTCCCCATGGCTGCCTGGGTCTCCATCTGTGTCCCTCTATGTCTCTCTGTGTTG  
CTTTTCAATTATGTCTCTTGGTATCTGGCTTGGGTTGTGTCTCTCCGTGTGACTATTTTGTCTCTCTCTCCCT  
CTCTTCTCTGTCTTCAGTCTCCGAATCTCCCCCTCTCTCTGTCTTCCCTGGTCCCTCTCTAGCCAGTGTGT  
CTCACCTGTATCTCTCTGCCAGGCTCTGTCTCTCGGTCTCTGTCTCACCTGTGCCTTCTCCCTACTGAACA  
CACGCACGGGATGGGCCTGGGGGGACCCTGAGAAAAGGAAGGGCTTTGGCTGGGCACGGTGGCTCACACCT  
GTAATCCAGCACTTTGGGAGGCCAAGGCAGGTAGATCACCTGAGGTGAGGAGTTCGAGACCAGCCTGGCC  
AACTGGTGAAACCCCATCTCTACTAAAAATACAAAAAATTAGCCAGGCGTGGTGGCGCATGCCTGTAGTCC  
CAGCTACTCAGGAGGCTGAGGGAGGAGAATTGCTTGAACCTGGGAGGTGGAGGTTGCAGTGAGCCGAGAC  
CGTGCCACTGCACTCCAGCCTGGGTGACAGAGTGAGACTCCGCCTCAAAAAAAAAAAAAAAAAAAAAAAAGAA  
AAGAAAAGAAAAGGAAGTGTTTTATCCCTGATGTGTGTGGGTATGAGGGTATGAGAGGGTCCCTCTCACT  
CCATTCCCTTCTCCAGGACATCCCTCCACTCTTGGGAGACACAGAGAAGGGCTGGTTCAGCTGGAGCTGGG  
AGGGGCAATTGAGGGAGGAGGAAGGAGAAGGGGAAGGAAACAGGGTATGGGGGAAAGGACCCTGGGG  
AGGGAAGTGAGGATACAACCTTGGGCCTGCAGGCCAGGCTACCTACCCACTTGGAACCCACGCCAAAGC  
CGCATCTACAGCTGAGCCACTCTGAGGCCTCCCCCTCCCCGGCGGTCCCCACTCAGCTCCAAAGTCTGTCTCC  
CTTTTCTCTCCACACTCTATCATCCCTGGATTCTCTCTACTTGGTTCTCATTCTTCTTTGACTTCCTG  
CTTCCCTTTCTCATTCTCTGTTTCTCACTTTCTGCCTGGTTTTGTCTTCTCTCTCTTTCTCTGGCGCA  
TGTCTGTTTCTCTATGTTTCTGTCTTTTCTTTCTCATCCTGTGTATTTTGGGCTCACCTTGTGTGCACTAT  
TCTCCCCTCTGCCCTTTTATTCTCTCTGTCTTTTACCTTCTTCCCTTTTCCCTTGGTTTTCTCTCAGTTTCT  
GTATCTGCCCTTACCCTCTCACACTGCTGTTTCCCAACTCGTTGTCTGTATTTTGGCCCGAACTGTGTCT  
TCCCCAACCTGTGTTTTTCTCACTGTTTCTTTTCTCTTTTGGAGCCTCCTCCTTGTCTCTGTCCCTTC  
TCTCTTTCATTATCATCCTCGCTCCTCATTCTGATCTGCTTCCCTCCCAGCAAAAGCGTGATCTTGCTGG  
GTCGGCACAGCCTGTTTCATCCTGAAGACACAGGCCAGGTATTTAGGTCAGCCACAGCTTCCCACACCCG  
CTCTACGATATGAGCCTCCTGAAGAATCGATTCTCAGGCCAGGTGATGACTCCAGCCACGACCTCATGCT

GCTCCGCCTGTCAGAGCCTGCCGAGATCACGGATGCTGTGAAGGTCATGGACCTGCCACCCAGGAGCCAG  
CACTGGGGACCACCTGCTATGCCTCAGGCTGGGGCAGCATTGAACCAGAGGAGTGTACGCCTGGGCCAGAT  
GGTGCAGCCGGGAGCCCAGATGCCTGGGTCTGAGGGAGGAGG**GGACAGGACTCCTGGGTCTGAGGGAGGA**  
**GGGCCAAGGAACCAGGTGGGGTCCAGCCCACAACAGTGTTTTTGCCTGGCCCGTAGTCTTGACCCCAAAGA**  
**AACTTCAGTGTGTGGACCTCCATGTTATTTCCAATGACGTGTGTGCGCAAGTTCACCCTCAGAAGGTGACC**  
**AAGTTCATGCTGTGTGCTGGACGCTGGACAGGGGGCAAAGCACCTGCTCGGTGAGTCATCCCTACTCCCA**  
**AGATCTTGAGGGGAAAGGTGAGTGGGAACCTTAATTCTGGGCTGGGGTCTAGAAGCCAACCAGGCGTCTG**  
**CCTCCCCTGCTCCCAGCTGTAGCCATGCCACCTCCCCGTGTCTCATCTCATTCCTCCTTCCCTCTTCTTTG**  
**ACTCCCTCAAGGCAATAGGTTATTCTTACAGCACAGCTCATCTGTTCTGCATTACAGCATACGGTTACTAG**  
**GCACCTGCTATGCACCCAGCACTGCCCTAGAGCCTGGGACATAGCAGTGAACAGACAGAGAGCAGCCCCCTC**  
**CCTTCTGTAGCCCCCATGCCAGTGAGGGGCACAGGCAGGAACAGGGACCACAACACAGAAAAGCTGGAGGG**  
**TGTCAGGAGGTGATCGGGCTCTGGGGGATGGAGAAGGGGTGGGGAGTGTGACTGGGAGGAGACATCCTGC**  
**AGAAGGTGGGAGTGAGCAAACACCTGCCGTAGGGGAGGGGAGGGGAGGGCCCTGCGGCACCTGGGGGAGC**  
**AGAGGGAACAGCATCTGGCCAGGCCTGGGAGGAGGGGCCTAGAGGGCGTGAGGAGCAGATTGGAGGTTGC**  
**CTGGCTGGAGTGAGGGATCGGGGCAGGGTGCGAGAGGGAAGAAAGGACCCCTCCTGCAGGGGCCTC**  
**ACCTGGGCCACAGGAGGACACTGCTTTTCCTCTGAGGAGTCAGGAGCTGTGGATG**  
**GTGCTGGACAGAAGCAGGACAGGGCCTGGCTCAGGTGTCCAGAGGCTGCCGCTGGCTTCCCTTTGGGATCA**  
**GACTGCAGGGAGGGAGGGCGGCAGGGGTGTGGGGGAGTGATGATGAGGATGACCTGGGGGTGGCTCCAG**  
**GCCTTGTCCCCGCCTGGGCCCTCACCCAGCCTCCCTCACAGTCTCCTGGCCCTCAGTCTCTCCCTCCACTCC**  
ATCCTCCATCTGGCCTCAGTGGGTCAATTCTGATCACTGAACTGACCGTACCCAGCCCTGCCATGGCCCTCC  
ATGGCTCCCCAATGCCCTGGAGAGGGGACATCTAGTCAGAGAGTAGTCCTGGAGAGGTGGCCTCTGCGATG  
TGCTGTGGGGGACAGATCCTGCAGATGGTCCCGGCCCTCATCCTGCTGACCTGTCTGCAGGGACTGTCTT  
CCTGGACCTTGCCCTTGTGCAGGAGCTGGACCCTGAAGTCCCTCCCATAGGCCAAGACTGGAGCCTTA  
TTCCCTCTGTTGGAATCCCTGCCATATTCTTGTGGGAGTG**A**GTTCCTGGAGACATTTCTGTCTGTTCTGA  
GAGCTGGGAATTGCTCTCAGTCATCTGCCTGCGCGGTTCTGAGAGATGGAGTTGCCTAGGCAGTTATTGGG  
GCCAATCTTTCTCACTGTGTCTCTCCTCCTTTACCTTAGGGTGATTCTGGGGGTCCACTTGTCTGTAATG  
GTGTGCTTCAAGGTATCACGTCATGGGGCCCTGAGCCATGTGCCCTGCCTGAAAAGCCTGCTGTGTACACC  
AAGGTGGTGCATTACCGGAAGTGGATCAAGGACACCATC**G**CAGCCAACCCCTGAGTGCCCTGTCCCACCC  
CTACCTCTAGTAAATTTAAGTCCACCTCACGTTCTGGCATCACTTGGCCTTTCTGGATGCTGGACACCTGA  
AGCTTGGAACCTCACCTTGCCAAAGCTCGAGCCTCCTGACTCCTACTGACCTGTGCTTTCTGGTGTGGAGTC  
CAGGGCTGCTAGGAAAAGGAATGGGCAGACACAGGTGTATGCCAATGTTTCTGAAATGGGTATAATTTTCG  
TCCTCTCCTTCGGAACAGTGGCTGTCTCTGGAGACTTCTCACTCAGTTTCAGTGAGGACACACACAAAGAC

GTGGGTGACCATGTTGTTTGTGGGGTGCAGAGATGGGAGGGGTGGGGCCCACCCTGGAAGAGTGGACAGT  
GACACAAGGTGGACACTCTCTACAGATCACTGAGGATAAGCTGGAGCCACAATGCATGAGGCACACACAC  
AGCAAGGATGACACTGAAAACATAGCCCACACTGTCCTGGGGGCACTGGGAAGCCTAGATAAAGGCCGTGA  
GCAGAAA<sub>g</sub>AAGGGGAGGGTCCTCCTATGTTGTTGAAGGAGGGACTAGGGGGAGAAACTGGCTGATTAATT  
ACAGGAGGTGTGTTTCAGGTCCCCCAAACCACCGTCACATTTGATGATTTCCCTAGCAGGACTTACAGAAATA  
AAGAGCTATCATGCTGTGGTTTATTATGGTTTCTTACATTGATAGGATACATACTGAAATCAGCAAACAA  
AACAGACGTATAGATTAGAGTGTGGAGAAAACAGAGGAAAACCTTGCAATTACGAAGACTAGCARCTTGGC  
TTTACTAAGTTTTTCAGACTGGCAGGAAGTCAAACCTATTAGGCTGAGGACCTTGTGGAGTGTAGCTGATC  
CAGCTGATAGAGGAACTAGCCAGGTGGGGGCCTTTCCCGGTGGGTGTGGGACATATCTGGCAAGATTTTG  
TGGCACTCCTGGTTACAGATACTGGGGCAGCAAATAAACTGAATCCTTGTTTTTCAGACCTTATACTGTGTA  
ATTGTAAGTGGACATTTAGTAATGTACAGGGGGTCAGGGAAAGACGGGGGATAGTGAATGCACCTGTGGG  
ACTGGTTTGGTTTTGTGATTGAGTAACCCATGGAGTTATCTCTGCAATGATGAATTGTTTGAAAGAAACC  
TGATTGTTTTTATATTGAATTAGAAGTTAATAAGCATGCTGAAAGATAAGAGAAGATGTTAGGAATATTG  
AGCAAGACTAACATACATAAAATATGAGAAAAAGTAGGTAGGAAATATTAGGTTGGAAAAATATGATAA  
TTGGAAAAAGTAAAAAAAAAATGGATTAATAGAATTGAGTAGAGGTGAAGAACGCAATAGTGACATTTT  
AGACCAGACTGAAGAAAAATGTCAAAGAAATACTGAATGATGTGAGTGCTGGCAGTGCAGATACTGAACA  
CTAAAGACTCAGGAAGAGAGAAGAATAGGAGAGGAGGAAACACATGACTGAACACTGAAGGCAGGTTTTT  
TTTTGTCTTTTTTTTTTTTTTTTTTTTTTTTCACAGAGTTTCACTCTTATCGCCCAGGCTGGAGTGCAATGGYT  
CAGTCTCGGCTCACTGCAACCCCTGCCTCCAGTTCAAGTGATTCTCCTGCTTCAGCCTCCCCAGTAGCTAG  
GATTACAGGTGCACACCACCACACCTGGCTAATTTTGCATTTTAAATAGAGACGGGGCTTCTCCATGTTGG  
TCAGGCTGGTCACGAGCTCCTGACCTCAGGTGATCCRCCTGCCTTGGCCTCCCAAAGTGCTGGGATTGATT  
ACAGGCGTGAGCCACCGCACCCAGCCTGAAGGTAAGTTTTATAGAATTTAAATACCTGAACGACCTCAGA  
TCAATAGAATCCATAGAGTGTGATTGAGTAGAAAGAAGGAAAGGCTATGAAACAAAAAAGAAAACTAGT  
ATTGTCATCGGCAAAAAATGTTTCATTGACAAAGTCCCTGGAGAGAAAGGATGGCGGAAATATTGAAGGAA  
AAGAATGTAGAACTGGAATATAATTCCTTCCAACAGTTTGGCAGTGTGTATCAATAGCCTTTTTTAGACA  
ATGCAACGTTCTCATACATTGCTTAAGGGAGTATAAATTGGTATCCTCATTTTAGAAAATGATTTAGCAG  
TATGTGTTGCAGCTGAACATATGCATGACCTATCAGTATCAATCTTAGAAATATACCCCTAGAAACCCTA  
GAAATACTTCATATTTTCATAAAAAATCCTCGTACCAGAAAGTTCATAGCATCCTCAAACCAGAACTGCCC  
AAAATGTTTATCTATAGGAGAGTGGAGAAATAAATGGCATATTCATACCAATGAATAGTGTGTGTAATGA  
ACAAGGATTATGTTTCATCCAAGACAGAGATGGGGCTTTAAaGAGACTGCCACAGGAACACAAAGAGAAA  
ACCAGAGACGGAGAGACAGCAGGGGAAATAGAATCTGAGATAGACAGTGACAAAATTGGGAAGAGAAAA

AACAAGGTAGCTAGAGACACACAGAGAGACACAGACAAACAGGAAGACAGCATCCCCTCGTGGTTAACAT  
CACAGCTTC

### Figure 5-3 Final Gorilla *KLK* Consensus

Area of sequence denoted in: yellow represents gaps in public assembly, green have additional nucleotides in the public assembly (A and AAA between existing highlighted GG) that do exist in my assembly, red represent synonymous nucleotide changes, in pink are nucleotides that exist in my assembly that do not exist in public assembly, in blue are SNP's, and in black are non-synonymous differences. The area indicated by all caps is the approximate region where the large deletion occurred to form the novel chimeric gorilla *KLK*.

**Table 5-1 ACPP Primers**

Primer Name	Sequence
ACPP_seqOrGo_Gap1	GTGGCACAATCTCGGCTC
ACPP_seqOrGo_Gap1rev	GCATGGTGAAACCCCATCTC
ACPP_seqOrGo_gap2	GTATTTCTTCCAGATTTT
ACPP_seqOrGo_gap2rev	GGATTACAGGCGTGCACCACC
ACPP_seqOrGo_gap3	GGTGGTGACGCCTGTAATCC
ACPP_seqOrGo_gap3rev	GCTGTTGATCAAGGTCACAGAGC
ACPP_seqOrGo_gap4	GAGTCTTGCTCTGTCACTCA
ACPP_X_ACC65I_Rev	GAGAAAAGGTACCTTACCAAAGTCACAACTTC
ACPP_X_HindIII_rev	GAGAAAAAGCTTTTACCAAAGTCACAACTTC
ACPP_SpliceOut1	GAAGTTTGTGACTTTCCTAAGTAGACTTTTCTC
ACPP_SpliceOut2	GAGAAAAGTCTACTTAGGAAAGTCACAACTTC
ATGtoAAGinACPP1	CAGCTCTCCTCAACAAGAGAGCTGCACCCCTCCTCCTGG
ATGtoAAGinACPP2	CCAGGAGGGGTGCAGCTCTCTTGTGAGGAGAGCTG
ACPP1370Fwd	CGCCTTTTGTAGTGTCCC
ACPP1350fwd2nd	GATCATAGCTGGAGAGGCAAG
ACPP1340fwd	GCAAGGATGTAACCTCACTC
ACPP1630fwd	ACC ACA TTC CAG GGC ATCC
ACPP1350fwd2nd	TAGCTGGAGAGGCAAGGATG
ACPPsryIntron380rev	GTCTTCATTAAGAGGAAGTTG
380rev_ACPP_HindIII	GTCTTCATTAAGAGGAAGTTGAAGCT

ATGtoAAGinACPP2	CCAGGAGGAGGGGTGCAGCTCTCTTGTGAGGAGAGCTG
ATGtoAAGinACPP1	CAGCTCTCCTCAACAAGAGAGCTGCACCCCTCCTCCTGG
ACPP_up_R5	TTAGTCACACCTTAAGACAA
ACPP_up_R4	AGGATGCCCTGGAATGTGG
ACPP_up_R1	CGAGAGAGGAAAGGCAAGAA
ACPP_up_F1	TGCAAGTTCCCTTGAATGGT
ACPP_up_F5	CAGCCACACCAGTCCAGAA
ACPP_up_F4	GAATCACTGGTCTAAGGTTTC
ACPP_up_F8	GTCCAATTGTCTTAAGGTGTG
ACPP_up_R6	TAGATAGGGTATGATACAGCGT
ACPP_up_R7	CTCTGTCAGGTTTGGAAAGG
ACPP_up_F7	CAGTTCTCCCATATACCACAT
ACPP_ATGgapR1	GTGCAGCTCTGTTGAGGAGAGCTG
ACPP_ATGgapF1	GCTCTCCTCAACAGAGCTGCACC
ACPP_pro338R1	GGTCTTCATTAAGAGGAACCTGC
ACPP_ATGgapR1	GGTGCAGCTCTGTTGAGGAGAGCTG
ACPP_ATGgapR2	CAGCTCTGTTGAGGAGAGCTGTTTCTG
ACPP_ATGgapR3	GGACGGGTGCAGCTCTGTTGAGGAG
ACPP_F9	CAACCTCCACCTCCTGGGTT
ACPP5primefwd	GGAGAGGCAAGGATGTAAC
ACPP5primerev	GTTACATCCTTGCCTCTCC
ACPPmidfwd	CCTCTATAGGCTACATGAGA
ACPPutrfd	GAAGTGTA TAGCCAAGGAG
ACPP_X_HindIII_rev	GAGTAAGCTTGAACAGAGCCTGGTCTCTGGGAGATGAAACCTGACGCTTG
ACPP_X_HindIII_w_splice_deldel	TGAGAAAAGCTTAGGAAAAGTCACAAACTTC
ACPP_normal_Cons_HindIII	GAGTAAGCTTGAACAGAGCCTGGTCTCTGGGAAACAATCCAAGGATAAG
ACPP_dupI_Cons_HindIII	GAGTAAGCTTGAACAGAGCCTGGTCTCTGGGAAACAATCCAAGGATAAAACAATCCAAGGATAAG
ACPP_Tripl_Cons_HindIII	GAGTAAGCTTGAACAGAGCCTGGTCTCTGGGAAACAATCCAAGGATAAAACAATCCAAGGATAAAACAATCCAAGGATAAG
ACPP_quad_Cons_HindIII	GAGTAAGCTTGAACAGAGCCTGGTCTCTGGGAAACAATCCAAGGATAAAACAATCCAAGGATAAAACAATCCAAGGATAAG

**Table 5-2 Gorilla BAC Hybrid Probes**

Probe Name	Sequence
upK3scott1Fprb	GTCCAGCCTCTTGAATTTCAAAGG
upK3scott1Rprb	CATCCAGCCTCTTCCTCCTTTGAA

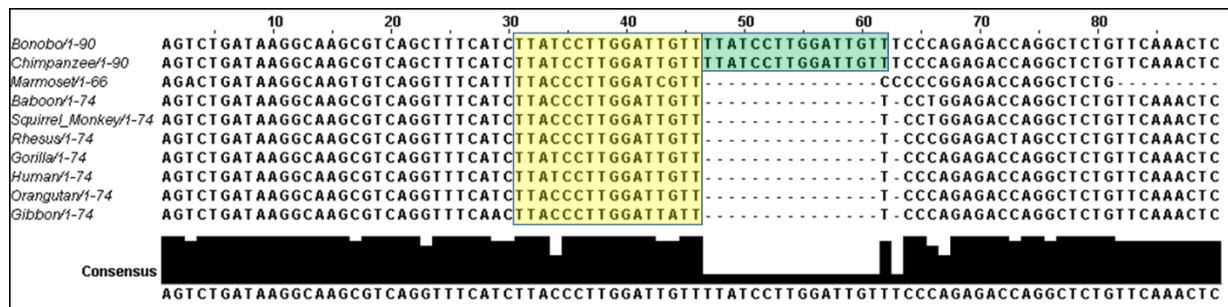
upK5scott2Fprb	CTGTTTCATCAGGCAGACACTTCTT
upK5scott2Rprb	GCAAGGAAGAAGAGATAAGAAGTG
upK1scott3Fprb	GTCCACTCATTCTGACCAAATCAC
upK1scott3Rprb	GAGAGCAGCCCACTCTGTGATTTG
KRSPscott4Fprb	CCTGGTTTACACTCCCCACAAAGG
KRSPscott4Rprb	GTAACCCCTGATCCCTCCTTTGTG
RPC194OVG_F	ATTCCCTAGGATGGTCTAGTCGCA
RPC194OVG_R	GAGGACCAAGAGAATTTGCGACTA

**Table 5-3 Gorilla Chimeric KLK primers.**

Primer Name	Sequence
GorNestrev2_6kb	ATTTCCGCCATCCTTTCTCT
GorNestrev1_2kb	TGGGTACGGTCAGTTCAGTG
GorNestfwd1	AGGTGTCTGTGTCCCAGAGG
GoPCRGapf1	CTGAACACTAAAGACTCAGG
GoPCRGapf2	CACATGACTGAACACTGAAGG
GoPCRGapR1	TGTAATCAATCCCAGCACTTTG
GoPCRGapR2	CACTCTATGGATTCTATTGATC
GoSeqf1	GAGTGTACGCCTGGGCCAGAT
GoSeqf2	GGTGCAGCCGGGAGCCCAGATG
GoSeqf3	GCCCTCAGTCTCTCCCCTCCAC
GoSeqf4	CTGTCTGTTCCCTGAGAGCTGG
GoSeq4	CCTGACTCCTACTGACCTGTG
GoSeqf6	GTTTCTTACATTGATAGGATAC
GoSeqR1	GTGATCAGAATGACCCACTG
GoSeqR2	GGAGGGGAGAGACTGAGGGCC
GoSeqR3	GAGACTGAGGGCCAGGAGAC
Gorilla_fwd_KRSP_GAP1	GAGTAGAAAGAAGGAAAGGC
Gorilla_rev_KRSP_GAP1	CTAGTTGAGCTCCTGCATGC
Gorilla_rev_KRSP_GAP2	CAGCCTGGGCGATAAGAGTG
Gorilla_fwd_KRSP_GAP2	GAAACACATGACTGAACACTG
Gorilla_fwd_KRSP_GAP2_b	GAAGGCAGGTTTTTTTTTTTGT
Gorilla_rev_KRSPpostgap2	CAAAAAAAAAACCTGCCTTC
Gorilla_rev_KRSPpostgap2_b	CTTCAGTGTTCAGTCATGTG
Gorilla_fwd_mid_Gap	GAGGGTATGAGAGGGTCCCTC
Gorilla_fwd_mid_Gap2	CCTGATCTAGCACCCCTCTGC
Gorilla_fwd_mid_Gap3	CTCACAGCTGCCCCACTGCATC
Gorilla_fwd_mid_Gap4	GTGTGACTATTTTGTCTCTC
KLK3fwd1feb	GTAAATCACCTGAGGTCAGG
KLK3forGofromHufwd	GAAAGGACCCTGGGGAGCGAAG
KLK3forGofromHu2fwd	CAGTTTCTGTATCTGCCCTTC
KLK3forGofromHu3fwd	GCATTGAACCAGAGGAGTGATC
KLK3forGofromHu4fwd	AGGAGGTGATCAGGCTCTCG
KLK3forGofromHu5fwd	GCATTGTCCCCACCTGGGCCC
KLK3forGofromHu6fwd	CTGCCTGTCCAGGTCTGAAAG
KLK3forGofromHu7fwd	GTGTCCTGGGGAATACTGGCC
KLK3forGofromHurev1	CTTCGCTCCCCAGGGTCCTTTC
KRSPrevs1feb	GGCTATTGATACACACTGCC
GoKLKJan17fwd1	CTGTCTCGGATTGTGGGAGG



GoKLKJan17fwd2	GCATTCCCAAACCCTGGCAG
GoKLKJan17fwd3	GTCTTTTCAAACCCACATCC
GoKLKJan17Rev1	GCTCCTGCATGCTGCCTCC
GoKLKJan17REV2	CTCTGCCTAGTTGAGCTCC
GoKLKJan17REV3	GAATGTATGATCGTGTTCC
KLK2ndfwd1	CCAAATCTGCATCCTTTTCAAAAC
KLK2ndfwd2	GCATCCCCAGCCACAATGC
KLK2ndfwd3	GATGTCTGATTCCCTAAATC
KRSP2REV1	CAATAATCTCCAAAACATCTGG
KRSP2REV2	GACTCAAGTTTATGAATTGAACG
KRSP2REV3	GATTCTTAAGTACACTAATTAC
GoKLK3firstintronfwd	CCAGCCCCACTCCAAGC
GoRevfromKRSP1	GGTCAATAGCTGCATATG
GoKLK3firstintronfwd_3	CAGTCCTCACTCCCACACC
GoKLK3firstintronfwd_2	CCAGCCCCTCCATATTGC
KRSP1seq_rev2	CAAAACATTAACCAGGTC
KRSP1seq_rev3	CATGGATAATAAAGCCTTGTC
revKLK2intron	CTGAAGTGTGGGCTGAAGGG
revKLK2intron2	GAGTCAGGGGCACATTGTCT
revKLK3intron1	TGGGGCTGGGAGAGACAGAA
revKLK3intron2	GGGCTGGGGGTATGGGCTTG
KLK1_F1	GGGGGCGGGTTCTGACTCTTATGC
KLK_R1	TTCTCCCCCTAGTCCCTCCTTCAACAAC
KLK_F2	TGAGGGGTTGAGTTGCAGGGGTTTGTTAGC
KLK_R2	CAGGGGTCGGGGAGATATAGTCGTGGTGAA
KLK_F3	TTGCGACTGTGGTTTGGGGTTTGGTATTTG
KLK_R3	CATTGTCTGGGCTGAAGTGTGGGCTGAAGG
KLK_F4	GGGGGTGAGGGAAAGGGAGAAGATGA
KLK_R4	AATACCAAACCCCAAACCACAGTCGCAACA
KLK2_gene_LR_F1	AGGGCAGGTCAGCAGGCATCTCT
KLK2_gene_LR_R1	TCCCTAGGCGGTTTCCTCATCTTCTC
KLK2_DN_LR_F1	GCCGCCCTCCTTCCCCCTTCC
KRSP1_LR_R1	TTCTCCTGCCCAACATGTTCTTCCTCTT
GoGAP_1fwd	TGCGAGAGGGAAGAAAGGAC
GoGAP_1rev	CAGACAGGTCAGCAGGATGA
GoGAP_2fwd	TTTGGGATCAGACTGCAGGG
GoGAP_2rev	CTCCAGAAGTCACTCCACACA



**Figure 5-4 Simian Alignment of conserved 16bp region in the first intron of ACPF**

The 16 base pair region, duplicated in chimpanzee is completely conserved in human and gorilla. This region has a C to T in the Great Apes, and is conserved through New World Monkeys.